Cena 15,00 zł
(VAT 8%)

Indeks 381306
e-ISSN 2543-8476
PL ISSN 0043-518X

WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

PAŹDZIERNIK / OCTOBER
ROCZNIK / VOLUME 69

2024 | 10

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

POLSKIE TOWARZYSTWO STATYSTYCZNE
POLISH STATISTICAL ASSOCIATION

GUS

TPS

WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

PAŹDZIERNIK / OCTOBER
ROCZNIK / VOLUME 69

2024 | 10 (761)

ADRES REDAKCJI I KONTAKT / EDITORIAL OFFICE ADDRESS AND CONTACT

sekretarz redakcji / Editorial Secretary: Małgorzata Zygmont (Główny Urząd Statystyczny, Polska)
Główny Urząd Statystyczny / Statistics Poland, al. Niepodległości 208, 00-925 Warszawa
ws.stat.gov.pl, e-mail: redakcja.ws@stat.gov.pl, tel./phone +48 22 608 32 25

Redakcja językowa: Wydział Czasopism Naukowych, Główny Urząd Statystyczny
Language editing: Scientific Journals Division, Statistics Poland

Redakcja techniczna, skład i łamanie, opracowanie materiałów graficznych i korekta:
Zakład Wydawnictw Statystycznych – zespół pod kierunkiem Macieja Adamowicza
Technical editing, typesetting, preparation of graphic materials and proofreading:
Statistical Publishing Establishment – team supervised by Maciej Adamowicz

Zakład Wydawnictw Statystycznych

Druk i oprawa / Printed and bound by:
Zakład Wydawnictw Statystycznych / Statistical Publishing Establishment
al. Niepodległości 208, 00-925 Warszawa, zws.stat.gov.pl

Wersja elektroniczna, stanowiąca wersję pierwotną czasopisma, jest dostępna na ws.stat.gov.pl
The primary version of the journal, issued in electronic form, is available at ws.stat.gov.pl

Informacje w sprawie sprzedaży i prenumeraty czasopisma / Sales and subscription of the journal:
Zakład Wydawnictw Statystycznych / Statistical Publishing Establishment
e-mail: zws-sprzedaz@stat.gov.pl, tel./phone +48 22 608 32 10, +48 22 608 38 10

Zam. 322/2024 – nakład 275 egz.

SPIS TREŚCI
CONTENTS

OD REDAKCJI

W październikowym numerze „Wiadomości Statystycznych. The Polish Statistician" znajdą Państwo cztery artykuły, w których przedstawiono wyniki badań przeprowadzonych z użyciem metod statystycznych, i opracowanie na temat zastosowania sztucznej inteligencji w statystyce publicznej.

Dr hab. Joanna Bruzda, prof. UMK, w pracy *Does modal (auto)regression produce credible forecasts of macroeconomic indicators?* bada przydatność regresji modalnej w analizie ekonomicznych jednowymiarowych szeregów czasowych na przykładzie indeksów produkcji przemysłowej 27 krajów OECD. Stawia pytanie, czy zastosowanie liniowych i nieliniowych (progowych) modeli autoregresji modalnej prowadzi do bardziej wiarygodnych prognoz makroekonomicznych niż w przypadku podejścia standardowego. Uzyskane wyniki wskazują, że liniowa autoregresja modalna może konkurować z innymi modelami wyspecyfikowanymi liniowo zarówno pod względem zagregowanych miar dokładności prognoz, jak i długości prognoz przedziałowych. Autorka zauważa przy tym, że węższe prognozy przedziałowe okazywały się często wysoce prawdopodobne w porównaniu z przewidywaniami uzyskanymi za pomocą innych metod estymacji, a różnice w miarach dokładności rzadko były istotne statystycznie. Wąskie prognozy przedziałowe wzrostu gospodarczego mogą raczej wymagać specyfikacji równania GARCH. Ponadto autoregresja modalna minimalizuje uodpornione wskaźniki sMAPE dla prognoz indeksów produkcji na jeden okres w przód. Autorka zastrzega jednak, że w innym kontekście empirycznym oraz przy zastosowaniu odmiennych schematów implementacji regresji modalnej i definicji funkcji straty rezultaty mogą się różnić od uzyskanych przez nią.

Dr hab. Marek Cierpiał-Wolan, prof. UR, dr Paloma Cuchí, dr Federico Gerardo de Cosio, mgr inż. Merkur Beqiri, dr Rifat Hossain i dr Dominik Rozkrut w artykule *Measuring and assessing the health of refugees from Ukraine in Poland: quantitative-qualitative mixed-methods approach* prezentują metodologię i kluczowe rezultaty kompleksowego badania „Stan zdrowia uchodźców z Ukrainy w Polsce" zrealizowanego wspólnie przez GUS i WHO. Obejmowało ono ankietę przeprowadzoną wśród członków gospodarstw domowych uchodźców ukraińskich po agresji Rosji z 24 lutego 2022 r., uzupełnioną badaniem behawioralnym opartym na pogłębionych wywiadach z uchodźcami. Ponadto przeprowadzono wywiady z polskimi i ukraińskimi przedstawicielami opieki zdrowotnej, organów odpowiedzialnych za działania ratunkowe i organizacji pozarządowych. Zastosowano metodę mieszaną, uwzględniającą badania ilościowe i jakościowe, a także wykorzystano dane z rejestrów administracyjnych i big data. Umożliwiło to przeprowadzenie analiz w czasie rzeczywistym i dokładniejsze zrozumienie potrzeb zdrowotnych uchodźców, ze szczególnym uwzględnieniem kobiet i osób starszych. Według autorów badanie uwidoczniło potrzebę pokonywania barier językowych i finansowych, jak również wynikających z nieznajomości systemu opieki zdrowotnej, związanych z ograniczeniami w dostępie do usług medycznych i barier kulturowych. Uniwersalność i elastyczność zastosowanej metody badawczej, zaaprobowanej przez Komisję Statystyczną ONZ, sprawia, że może ona być stosowana również w innych krajach zmagających się z kryzysami uchodźczymi.

W pracy *Czy warto budować modele prognozowania upadłości przedsiębiorstw uwzględniające kryterium przynależności sektorowej?* dr Sergiusz Herman i dr Bartłomiej Lach weryfikują zasadność konstruowania sektorowych modeli prognozowania upadłości przedsiębiorstw. Analizę przeprowadzają na podstawie danych finansowych dotyczących 800 polskich przedsiębiorstw za lata 2015–2019 (pobranych z bazy EMIS Professional), stosując sześć metod doboru zmiennych, sześć metod uczenia maszynowego oraz 500 różnych losowych prób uczących i testowych. Dochodzą do wniosku, że

w poszczególnych sektorach gospodarki istnieją odmienne determinanty upadłości. Modele opierające się na grupie przedsiębiorstw prowadzących jednorodną działalność gospodarczą pozwalają uzyskać przeciętnie wyższe wskaźniki trafności klasyfikacji. Autorzy zauważają także, że trafność klasyfikacji modelu prognostycznego jest uzależniona od wielkości próby badawczej, a modele konstruowane z wykorzystaniem tych samych metod prognostycznych klasyfikują badane obiekty w podobny sposób.

Artykuł dr. hab. Krzysztofa Najmana, prof. UG, dr hab. Kamili Migdał-Najman, prof. UG, mgr Katarzyny Racy i mgr Agaty Majkowskiej *Identifying age groups of Twitter users based on the specific characteristics of textposts* jest poświęcony możliwości określenia grup wieku autorów wpisów w serwisie Twitter (obecnie X) na podstawie elementów zwykle usuwanych z tekstów analizowanych metodami text mining, takich jak emotikony, znaki interpunkcyjne i słowa, które nie są nośnikami treści (ang. *stopwords*). Badacze analizują prawie 3 mln tweetów w języku angielskim opublikowanych przed lipcem 2020 r. Z przeprowadzonego badania wynika, że wyodrębnione cechy w niewielkim stopniu różnicują grupy wiekowe. Najbardziej specyficznym stylem pisania wiadomości wyróżniają się najmłodsi użytkownicy Internetu.

The role of Artificial Intelligence when generating official statistical data to temat podjęty przez dr Arzu Taghiyevą. Autorka zauważa, że dzięki stosowaniu rozwiązań opartych na sztucznej inteligencji (ang. *artificial intelligence* – AI) – zwłaszcza w gromadzeniu danych z różnych źródeł i dostosowywaniu się do ich specyfiki, przetwarzaniu danych i kontroli jakości oraz eksploracji nowych metod, a także udostępnianiu danych – instytucje statystyczne mogą poprawić wydajność, dokładność i terminowość procesów produkcji statystycznej. Duże wyzwanie w korzystaniu z AI stanowią jednak m.in. kwestie etyczne i prawne oraz ochrona prywatności.

Numer zamyka prezentacja najnowszych publikacji GUS przygotowana przez Joannę Sadowy.

Życzymy miłej lektury.

FROM THE EDITORIAL TEAM

In the October issue of *Wiadomości Statystyczne. The Polish Statistician* there are four articles presenting the results of research carried out by means of statistical methods, and a work on the application of artificial intelligence in official statistics.

Joanna Bruzda, PhD, DSc, professor at Nicolaus Copernicus University in Toruń, in her paper *Does modal (auto)regression produce credible forecasts of macroeconomic indicators?* assesses the usefulness of modal regression for the analysis of univariate economic time series on the example of industrial output indexes of 27 OECD countries. The author poses a question whether the application of linear and nonlinear (threshold type) models of modal autoregression leads to more credible macroeconomic forecasts than in the case of the standard approach. The obtained results show that linear modal autoregression can compete with other linearly-specified models both in terms of the aggregate forecast accuracy and the length of prediction intervals. The author also observes that narrower prediction intervals in comparison with other estimation methods often turn out highly likely, and differences in precision measurements rarely prove statistically significant. Narrow-interval forecasts of economic growth rates might rather require specifying a GARCH equation. Moreover, modal autoregression minimises the robustified sMAPE indicators for one-step-ahead forecasts of output indexes. The author admits, however, that in a different empirical context and using different schemas of modal autoregression implementation and different definition of the loss function, similar research might yield disparate results.

In *Measuring and assessing the health of refugees from Ukraine in Poland: quantitative-qualitative mixed-methods approach* by Marek Cierpiał-Wolan, PhD, DSc, Professor at Rzeszów University, Paloma Cuchí, MD, MPH, Federico Gerardo de Cosio, MD, MPH, Merkur Beqiri, BEng, MSc, Rifat Hossain, PhD, and Dominik Rozkrut, PhD, the authors present the methodology and key findings

of the 'Health of Refugees from Ukraine in Poland' survey carried out jointly by Statistics Poland and the WHO. The survey included questionnaires conducted among members of refugee households following the Russian agression on Ukraine on 24th February 2022, supplemented by behavioural insights research based on in-depth interviews with those refugees. Additionally, interviews were conducted with Polish and Ukrainian healthcare providers, authorities responsible for rescue operations and representatives of non-governmental organisations. The authors, in addition to using data from administrative registers and big data, applied a mixed-methods approach that takes into account both quantitative and qualitative research. This made it possible to carry out real-time analyses and understand the refugees' healthcare needs in more detail, with particular regard to the needs of women and older people. According to the authors, their study highlights the importance of addressing language and financial barriers, issues related to the unfamiliarity with the health system, as well as healthcare service constraints and cultural barriers. The universal character and flexibility of this approach and the fact that it was approved by the UN Statistical Commission renders it suitable for application in other countries facing a refugee crisis.

The article *Is it worth building models for forecasting bankruptcy of enterprises taking into account the criterion of sectoral affiliation?* by Sergiusz Herman, PhD, and Bartłomiej Lach, PhD, attempts to verify the legitimacy of constructing industry models for forecasting the bankruptcy of enterprises. The research is based on financial data of 800 Polish companies for the years 2015–2019, drawn from the EMIS Professional database. The authors use six methods of variables selection, six machine learning methods and 500 distinct random training and testing samples. They formulate a conclusion that different sectors of the economy have different determinants of bankruptcy. Models constructed on the basis of a group of enterprises conducting homogeneous economic activity make it possible to achieve, on average, higher classification accuracy rates. Another conclusion is that the accuracy of the classification of a forecasting model depends on the size of the research sample, and various models constructed by means of the same prognostic methods classify the studied objects in a similar way.

The article by Krzysztof Najman, PhD, DSc, professor at the University of Gdańsk, Kamila Migdał-Najman, PhD, DSc, professor at the University of Gdańsk, Katarzyna Raca, MSc, and Agata Majkowska, MSc, *Identifying age groups of Twitter users based on the specific characteristics of textposts* focuses on the possibility of determining age groups among the authors of texts posted on Twitter (now X) based on elements usually removed from analysed texts by means of text mining methods. These elements include: emoticons, punctuation marks and words that do not carry content (stopwords). The researchers analyse nearly 3 million tweets in English published before July 2020. Their study shows that the distinguished characteristics differentiate user age groups only to a small extent. The youngest Internet users have the most specific style of writing textposts.

In the course of her study presented in the article entitled *The role of Artificial Intelligence when generating official statistical data*, Arzu Tagaieva, PhD, notes that by using solutions based on artificial intelligence (AI), especially in collecting data from different sources and adapting to their specificity, processing data and quality control as well as exploring new methods and sharing data, statistical institutions can improve the efficiency, accuracy and timeliness of statistical production processes. However, the idea of privacy protection as well as ethical and legal issues pose a serious challenge in using AI.

The issue concludes with the presentation of Statistics Poland's most recent publications prepared by Joanna Sadowy.

We wish you pleasant reading.

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(10), 1–27 https://doi.org/10.59139/ws.2024.10.1
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 29.05.2023, poprawiony/revised: 1.07.2024, zaakceptowany/accepted: 2.09.2024

Does modal (auto)regression produce credible forecasts of macroeconomic indicators?[1]

Joanna Bruzda[a]

Abstract. Modal regression is a relatively new statistical technique that serves the purpose of modelling the modal value of a variable conditional on a set of explanatory variables. The aim of the study discussed in this paper is to assess the usefulness of modal regression in the analysis of economic time series and, more specifically, in short-term macroeconomic forecasting, on the example of industrial output indexes of 27 OECD countries. We focus on models of univariate time series and pose a question whether linear and nonlinear (threshold type) modal autoregression leads to more trustworthy macroeconomic forecasts than standard approaches based on modelling other measures of central tendency. The credibility of forecasts is understood as appropriately defined accuracy of point forecasts, and as narrow forecast intervals.

Our empirical results demonstrate that linear modal autoregression can compete with other linearly-specified models both in terms of the aggregate forecast accuracy and the lengths of prediction intervals. It should be mentioned, however, that in the case of the study described in this paper, carried out on the data sample spanning the period from January 1994 to December 2019 and the period of forecasting allowing the determination of 100 forecasts in time horizon from one to four months according to the rolling schema, narrower prediction intervals in comparison with other estimation methods have turned out more likely for lower nominal confidence levels such as 80% and below, while differences in precision measurements, including credibility differences, have rarely proved statistically significant. Our analysis also shows that narrow-interval forecasts of economic growth rates might require specifying a GARCH equation, but on the other hand, at certain confidence levels and forecast horizons, models with GARCH equations might be outperformed in this respect by nonlinear (in mean or mode) autoregression. Another interesting fact is that modal autoregression minimises the robustified sMAPE indicators for one-step-ahead forecasts of the output indexes.

Keywords: forecasting, modal regression, industrial production, univariate time series models
JEL: C13, C22, C53, E37

Czy (auto)regresja modalna dostarcza wiarygodnych prognoz wartości wskaźników makroekonomicznych?

Streszczenie. Regresja modalna jest stosunkowo nowym narzędziem statystycznym, służącym do opisywania wartości modalnej zmiennej zależnej warunkowo względem zbioru ustalonych zmiennych objaśniających. Celem badania omawianego w artykule jest określenie przydatności

[1] Artykuł został opracowany na podstawie referatu wygłoszonego na XVII Ogólnopolskim Seminarium Naukowym Profesora Zygmunta Zielińskiego „Dynamiczne modele ekonometryczne", które odbyło się w dniach 6–7 września 2021 r. w Toruniu. / The article is based on a paper delivered at the 17th Professor Zygmunt Zieliński International Conference on Dynamic Econometric Models, held on 6–7 September 2021 in Toruń, Poland.

[a] Uniwersytet Mikołaja Kopernika w Toruniu, Wydział Nauk Ekonomicznych i Zarządzania, Polska / Nicolaus Copernicus University in Toruń, Faculty of Economic Sciences and Management, Poland.
ORCID: https://orcid.org/0000-0002-7451-3592. E-mail: joanna.bruzda@uni.torun.pl.

regresji modalnej w analizie ekonomicznych szeregów czasowych, a dokładniej – w krótkookresowym prognozowaniu makroekonomicznym, na przykładzie indeksów produkcji przemysłowej z 27 krajów OECD. Skupiając się na modelach jednowymiarowych szeregów czasowych, postawiono pytanie, czy zastosowanie liniowych i nieliniowych (progowych) modeli autoregresji modalnej prowadzi do uzyskania bardziej wiarygodnych prognoz makroekonomicznych niż w przypadku podejścia standardowego, polegającego na modelowaniu innych miar tendencji centralnej. Wiarygodność prognoz jest tu rozumiana jako odpowiednio zdefiniowana dokładność prognoz punktowych oraz ograniczona szerokość prognoz przedziałowych.

Wyniki analizy empirycznej wskazują, że liniowa autoregresja modalna może konkurować z innymi modelami wyspecyfikowanymi liniowo zarówno pod względem zagregowanych miar dokładności prognoz, jak i długości prognoz przedziałowych. Należy jednak zauważyć, że w przypadku badania omawianego w artykule, przeprowadzonego na próbie z okresu od stycznia 1994 r. do grudnia 2019 r., z wydzielonym okresem prognozowania pozwalającym na wyznaczenie 100 prognoz w horyzoncie od jednego do czterech miesięcy zgodnie ze schematem rolling, węższe prognozy przedziałowe w porównaniu z innymi metodami estymacji okazały się bardziej prawdopodobne dla poziomu ufności 80% i poniżej, a różnice w miarach dokładności, w tym różnice w wiarygodności, były rzadko istotne statystycznie. Z analizy wynika również, że wąskie prognozy przedziałowe wzrostu gospodarczego mogą raczej wymagać specyfikacji równania GARCH, choć dla pewnych poziomów ufności i horyzontów prognoz modele z równaniami GARCH mogą być pod tym względem zdominowane przez modele nieliniowej warunkowej średniej lub warunkowej mody. Interesującym wnioskiem jest to, że autoregresja modalna minimalizuje uodpornione wskaźniki sMAPE dla prognoz indeksów produkcji na jeden okres w przód.

Słowa kluczowe: prognozowanie, regresja modalna, produkcja przemysłowa, modele jednowymiarowych szeregów czasowych

1. Introduction

1.1. Motivation and research objectives

As noted by Manski (1991), the meaning of the term 'regression' has evolved to encompass many different statistical concepts. Throughout this paper, we will apply his definition of regression:

> A regression of y on x is any feature of the probability distribution of y conditional on x, considered as a function of x. The feature of interest might be the mean, median, mode, or variance. Hence, we speak of the mean regression, median regression, mode regression, variance regression, and so on (Manski, 1991, p. 34).

Modal regression can be motivated in many different ways. The interpretation of the parameters of a modal regression line as showing a reaction of a 'typical' object of interest, being it an enterprise, a client, or a patient responding to a certain economic instrument, promotional activity, or medical treatment, makes it an interesting alternative to the traditional mean regression. Kemp and Santos Silva (2012) note that the (conditional) mode is the most intuitive measure of central tendency. They also cite articles by Temple (e.g. 1999) who argues that, in the case of the growth regression based on cross-sectional and panel data, one should use robust estimation techniques

enabling the fitting of a regression line to the majority of the data. We feel that this statement is even more true in the case of modelling time series of growth rates, especially taking into consideration the global financial crisis of 2007–2008 and the more recent turmoil in the world economy caused by the COVID-19 pandemic and the Russian invasion of Ukraine. All these events resulted in outlying values of at least certain growth rates.[2]

On the other hand, Dimitriadis et al. (2024) discuss the notion of mode or modal forecast and cite a paper by Kröger and Pierrot (2019), where the authors show that mode is the measure of a central tendency most frequently chose by participants of laboratory experiments asked to issue a point forecast. The authors also indicate that individual income survey forecasts are rationalised in the best way as mode forecasts. In addition, central banks often report mode forecasts of inflation and growth rates to show the most likely economic inflation scenario and the most likely trajectories of growth (the 'central scenario' – Pońsko & Rybarczyk, 2016). In theory, such forecasts might be considered as the most trustworthy (credible) assuming, as in e.g. Gambetta (1988, pp. 217–218), that trust is understood as 'a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action' and 'is better seen as a threshold point, located on a probabilistic distribution of more general expectations, which can take a number of values suspended between complete distrust (0) and complete trust (1), and which is centered around a mid-point (0.50) of uncertainty'.

It must be noticed, however, that modal forecasts of absolutely continuous random variables, understood as the values maximising the corresponding probability density function, are not elicitable, i.e., they cannot be obtained through the minimization of the expected value of a certain loss function, when considering both the classes of unimodal distributions and strongly unimodal distributions (i.e. unimodal distributions with a unique local mode; Heinrich, 2014; Heinrich-Mertsching & Fissler, 2021). What is elicitable is the mid-point of the modal interval (e.g. Gneiting, 2011). For this reason, Dimitriadis et al. (2024) introduce the notion of asymptotic elicitability, meaning that there exists a sequence of elicitable functionals converging to the mode, and show that the mode is asymptotically elicitable in the class of absolutely continuous and unimodal (strongly unimodal) distributions. This enables the authors to suggest an asymptotically-valid procedure for testing the rationality of modal forecasts. In this

[2] Another potentially interesting area of application of modal regression is statistical quality management, where quality is often assessed with a step loss function incorporated in process capability indices or, more generally, in designing managerial instruments under the assumption of a bounded forecast cost function (compare Kemp et al., 2020. The authors assert that the motivation for using the step loss function also comes from statistical process control). In addition, investors in financial markets can potentially benefit from modal regression if, for example, they use complex option strategies and forecast the price of the underlying asset.

paper, we supplement the evaluation of modal forecasts with a family of forecast accuracy measures which can be understood as measuring the trustworthiness of forecasts and which are based on symmetric and bounded loss functions, including those used in modal regression estimation.

Modal regression is a direct approach to compute modal forecasts, enabling one to perform the necessary computations in one step. This approach might be interpreted as a loss-based estimation of the forecasting model (for the introduction to loss-based estimation; see Elliott & Timmermann, 2016). As noted by Elliott and Timmermann (2008), forecast quality might often be improved by using the same loss function in the estimation and forecasting, which is documented in the literature (e.g. Bruzda, 2019; Weiss, 1996). In our work, we pose a question whether the credibility of macroeconomic forecasts can be boosted through the use of modal regression combined with univariate-time-series models, which leads to linear and nonlinear modal autoregression. The gist of this question can be expressed in two separate ones, the first of which relates to simple ex-post forecast accuracy measures assessing forecast credibility, while the second concerns the lengths of prediction intervals, which in the case of applying modal regression are expected to be narrower (e.g. Chen et al., 2016; Xiang & Yao, 2022; Yao & Li, 2014). In addition to higher trustworthiness of forecasts, it can also be expected that modal regression will robustify the forecasting outcomes with regard to problems such as outliers, conditional skewness, heavy tails and nonlinearities taking the form of, e.g., structural parameters changing across quantiles and conditional heteroscedasticity of a general form. Among the economic variables whose nonlinearity is often confirmed in empirical studies, there are output measures and unemployment rates (e.g. Granger & Teräsvirta, 1993, and their references to the literature). Besides empirical observations, there are also good theoretical explanations for the observed nonlinearity of macroeconomic variables over the business cycle (see, e.g., the discussion and references in Bruzda, 2007). Moreover, it is documented that growth rates are skewed and leptokurtic, often even when heteroscedasticity in the data has been removed (Berger et al., 2020; Fagiolo et al., 2008; Stockhammar & Öller, 2011; Ul Hassan & Stockhammar, 2016). Due to the above, we believe that forecasts of macroeconomic indicators, including output indexes, can potentially be improved by means of modal regression.

1.2. Brief history of modal regression

The concept of modal regression has been around for decades, with the first theoretical results dating back to Sager and Thisted (1982), who studied a specific nonparametric modal regression under shape constraints (the isotonic modal regression), and Lee (1989), who focused on the case of a loss function based on the

uniform (rectangular) kernel of a fixed length and examined statistical properties of the resulting regression. The latter is also known as δ-mode regression (Manski, 1991), assuming that the conditional distribution of the regressand is unimodal and symmetric in the modal interval of length 2δ. The above-mentioned research was inspired by the observation that, under certain assumptions, the conditional mode of truncated data coincides with the conditional mean, and thus it can be classified as one concerning robust M-estimation of the mean regression line. It was not until the publications of Kemp and Santos Silva (2012) and Yao and Li (2014) that the statistical apparatus of semiparametric modal regression per se was developed, i.e., the consistent estimation of the conditional mode line was introduced. More recently, Feng et al. (2020) observed that a regression problem based on maximising the correntropy criterion with the parameter tending to zero scale is essentially a modal regression. Correntropy is a generalised dependence measure gaining popularity in the statistical learning literature due to its robustness properties.

The papers by Kemp and Santos Silva (2012) and Yao and Li (2014) deal with linear modal regression estimated with *iid* data and also consider some implementation issues, such as a grid search over possible values of the bandwidth parameter (Kemp & Santos Silva, 2012) and a bandwidth minimising the asymptotic weighted mean squared error (Yao & Li, 2014). Subsequently, the idea of modal regression has also been studied in the context of, for example, partially linear regression (Krief, 2017; Ullah et al., 2023), panel data (Ullah et al., 2021), linear regression with autocorrelated errors (Wang, n.d.) and nonparametric regression (Chen et al., 2016; Einbeck & Tutz, 2006; Xiang & Yao, 2022; Yao et al., 2012). In the latter studies, it is often assumed that the conditional distribution can be multimodal, and it is expected that nonparametric modal regression will be effective at capturing multiple local trends in the response variable. Einbeck and Tutz (2006) provide an example of an empirical study taking advantage of this property of modal regression as applied to a transportation problem.

The concept of modal regression can also be applied to stationary time series data on condition that certain memory and moment requirements are introduced. A formal statistical framework for dealing with this case was developed by Kemp et al. (2020) and Ullah et al. (2022). In the former of these studies, the authors consider the estimation of vector-autoregressive-conditional-mode models describing the conditional mode of a random vector having a well-defined global mode and a continuous conditional joint density. They argue that the difference between the conditional modal value of a random vector and the vector of marginal modal values can be substantial. As a byproduct, they generalise the previous results on linear modal regression to the case of time series. They are also the first researchers who consider a stochastic bandwidth. In contrast, the work by Ullah et al. (2022) focuses on nonlinear modal regression with dependent data. The authors show that in this case,

the asymptotic distribution of the modal estimator and the optimal bandwidth can be expressed in the same way as for linear models based on *iid* data. This is due to the fact that in an asymptotic analysis, the condition imposed on the bandwidth parameter makes it possible to treat error terms as if they were independent. Nonlinear-autoregressive-time-series models are of special interest for their research.

1.3. Research strategy

The above-mentioned results are directly employed in our study, which, as mentioned before, is devoted to the empirical verification of the predictive properties of a modal regression in forecasting industrial output in 27 OECD countries with linear and nonlinear autoregressive models. For example, Chauvet and Potter (2013) assert that linear and nonlinear autoregressions had the best accuracy in forecasting the U.S.'s (GDP-measured) output growth in real time during normal times (i.e. expansions) when compared with the current depth of recession models, dynamic-stochastic-general-equilibrium models, (Bayesian) vector autoregressions, dynamic-factor models with Markov switching, and judgmental forecasts. We use data from the period of January 1994 to December 2019 and compute point and interval forecasts from one to four steps ahead in a rolling forecasting schema resulting in 100 forecasts for each forecast horizon. This constitutes our contribution to the ongoing debate on the usefulness of modal regression in time series analysis, or, more precisely, in short-term macroeconomic forecasting. To the best of our knowledge, such research has not been so far undertaken in the economic context. The discussion of the research outcomes is preceded by a detailed presentation of the methodology, which – as regards modal regression and its uses in forecasting – is not widely known.

2. Modal regression

The model considered in this paper has the following form:

$$Mode(y_t|\Omega_{t-1}) = g(y_{t-1}, y_{t-2}, \dots, y_{t-k}; \boldsymbol{\beta}), \tag{1}$$

where $y_t, t = 1, 2, \ldots$, is a vector of observations, Ω_{t-1} is the σ-algebra generated by past observations, g is a certain function of past observations on y_t, and $\boldsymbol{\beta}$ is the vector of parameters.

It is assumed that the conditional density of $\varepsilon_t = y_t - g(y_{t-1}, y_{t-2}, \dots, y_{t-k}; \boldsymbol{\beta})$ has a well-defined global mode at zero (Kemp & Santos Silva, 2012; Ullah et al., 2021). To estimate the parameters, one can solve the optimisation problem formulated as

$$\sum_t \Phi_h\big(y_t - g(y_{t-1}, y_{t-2}, \ldots, y_{t-k}; \boldsymbol{\beta})\big) \to \max, \tag{2}$$

where $\Phi_h(u) = \frac{1}{h}\Phi\left(\frac{u}{h}\right)$ and $\Phi(u)$ is a non-negative kernel function symmetric around 0, and h is the bandwidth parameter.

Although, in principle, any well-behaved kernel function can be used, the Gaussian kernel has been adopted by default (compare Kemp et al., 2020; Kemp & Santos Silva, 2012; Ullah et al., 2022; Yao & Li, 2014). The main reason for this choice is the fact that, with different values of h, the estimation based on (2) nests as limiting cases both the estimation of the conditional mean and the conditional mode. Furthermore, this kernel function leads to a computation algorithm similar to the iterated weighted least squares.

It is worth mentioning that solving the optimisation problem (2) assuming the Gaussian kernel is equivalent to the (non-studentised) robust M-estimation with the Welsch weight function. However, under a judicious choice of bandwidth, namely taking $h \to 0$ and $h^5 n \to \infty$ and assuming certain regularity conditions, it is possible to obtain consistency results with respect to the parameters of the conditional mode equation, even if the error term is asymmetric (Ullah et al., 2022). The asymptotic normality result requires a further tightening of the requirements regarding h and, in the case of models linear in parameters, is formulated as follows (Yao & Li, 2014):

$$\sqrt{nh^3}\left[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} - \frac{h^2}{2}J^{-1}M\{1 + o(1)\}\right] \Rightarrow N\{0,\ \nu_2 J^{-1}LJ^{-1}\},$$

where $\nu_2 = \int t^2\,\Phi^2(t)dt$, $J = E[f''(0|\boldsymbol{x}_t)\boldsymbol{x}_t\boldsymbol{x}_t^{\mathrm{T}}]$, $M = E[f'''(0|\boldsymbol{x}_t)\boldsymbol{x}_t]$, $L = E[f(0|\boldsymbol{x}_t)\boldsymbol{x}_t\boldsymbol{x}_t^{\mathrm{T}}]$, $f(\cdot)$ denotes the probability-density function of the error term, and $\boldsymbol{x}_t$ is the column vector of regressors.

Yao and Li (2014) also introduced a numerical solution to modal estimation called the Modal Expectation-Maximization (MEM) algorithm, which, in the case of the Gaussian kernel, is equivalent to the iteratively reweighted least squares. Assuming that models linear in parameters are estimated, which is the case in our paper, the MEM algorithm is the following (Yao & Li, 2014). Starting with a good robust initial estimate of $\boldsymbol{\beta}$, denoted $\widehat{\boldsymbol{\beta}}^{(0)}$, repeat until convergence:

E-step: calculate the weights:

$$\pi\big(t|\widehat{\boldsymbol{\beta}}^{(k)}\big) = \frac{\Phi_h\big(y_t - \boldsymbol{x}_t^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^{(k)}\big)}{\sum_t \Phi_h\big(y_t - \boldsymbol{x}_t^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^{(k)}\big)}.$$

M-step: update the estimates:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \underset{\widehat{\boldsymbol{\beta}}}{\operatorname{argmax}} \sum_t \left\{\pi\left(t|\widehat{\boldsymbol{\beta}}^{(k)}\right) \log \Phi_h\left(y_t - \boldsymbol{x}_t^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\right)\right\} = (\mathbf{X}^{\mathrm{T}}\mathbf{W}_k\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}_k\mathbf{Y},$$

where $\mathbf{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^{\mathrm{T}}$, $\mathbf{Y} = (\boldsymbol{y}_1, \mathbf{y}_2, \ldots, \boldsymbol{y}_n)^{\mathrm{T}}$, and

$$\mathbf{W}_k = \begin{pmatrix} \pi\left(1|\widehat{\boldsymbol{\beta}}^{(k)}\right) & & \mathbf{0} \\ & \pi\left(2|\widehat{\boldsymbol{\beta}}^{(k)}\right) & \\ & & \ddots \\ \mathbf{0} & & \pi\left(n|\widehat{\boldsymbol{\beta}}^{(k)}\right) \end{pmatrix}.$$

The initial estimate in our computations is the one obtained with robust linear regression based on Huber's function. Note that Yao and Li (2014) advise using multiple initial estimates.

Yao and Li (2014) and Ullah et al. (2022) also derive an asymptotically optimal value of parameter h, i.e., the value minimising the asymptotic mean squared error. Assuming that the weight matrix reflecting which coefficients are more important in inference is proportional to the inverse of the asymptotic variance of $\widehat{\boldsymbol{\beta}}$, the optimal bandwidth in the linear case is the following:

$$h_{opt} = \left[\frac{3p\int t^2\,\Phi^2(t)dt}{M^T L^{-1} M}\right]^{\frac{1}{7}} n^{-\frac{1}{7}},$$

where p is the number of parameters. This particular bandwidth parameter is adopted in this paper and is found using a plug-in estimation.

Among the potential benefits of using modal regression, the following could be mentioned:

- modal regression makes it possible to measure marginal effects for the largest fractions of units in the examined population, being robust to the skewness of the conditional distribution. In a similar way, it enables a researcher to find the most common reaction function of a single examined unit over time;
- it is robust to outlying observations and tick tails;
- it allows a direct computation of modal forecasts;
- in the case of strongly unimodal error distributions, it can produce modal forecasts with higher values of the related accuracy (i.e. credibility) measures than, e.g., mean or median regression;
- in the case of strongly unimodal error distributions, it can help to find narrower prediction intervals than, e.g., mean or median regression (Chen et al., 2016; Yao & Li, 2014);

- it can lead to better estimates of regression coefficients in the conditions of asymmetry of the error distribution (Yao & Li, 2014).

To illustrate how modal regression is related to its two main competitors – mean and median regression – we generated 500 random samples from the following model:

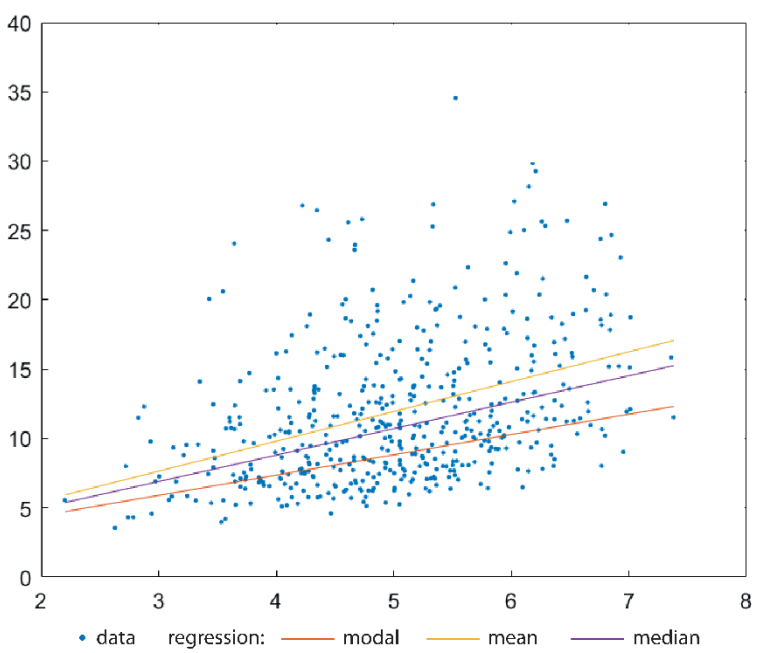
$$Y = X + (1 + 0.5X)\varepsilon,$$

where $X \sim N(5, 1)$ and $\varepsilon \sim \Gamma(2, 1)$, and estimated regression lines using all three approaches.

The results are presented in Figure, which shows that the modal regression line goes through the highest density region, which may result in narrower prediction intervals around the regression line than those generated with the help of mean or median regression.

The above-mentioned benefits of this relatively new statistical tool appear to be substantial and call for detailed evaluations in empirical settings. One such evaluation is undertaken here in Section 4 and focuses on macroeconomic forecasting.

Figure. Modal, mean, and median regression lines for simulated data



Source: author's computations.

3. Prediction intervals and forecast evaluation

Yao and Li (2014) presented a method of computing asymmetric prediction intervals around the conditional mode, which applies if the distribution of the error term in the modal regression model is independent of the regressors and unimodal. The method is the following: assuming that $\widehat{\boldsymbol{\beta}}$ is the vector of estimated coefficients of the modal regression equation and that the residuals $e_t = y_t - \boldsymbol{x}_t^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$ from this regression have also been computed, we denote by $e_{[i]}$ the i-th smallest residual ($i = 1, \dots, n$). Let $\hat{f}$ be the kernel density estimate of f based on the residuals e_t. The idea is to find such $k_1 < k_2$ that for a given confidence level $1-\alpha$, $k_2 - k_1 = [n(1-\alpha)]$ and $\hat{f}(e_{[k_1]}) \approx \hat{f}(e_{[k_2]})$. To find indexes k_1 and k_2, Yao and Li (2014) suggest taking $k_1 = \left[\frac{n\alpha}{2}\right]$ and $k_2 = n - k_1$, and repeating the following two steps:

Step 1:
if $\hat{f}(e_{[k_1]}) < \hat{f}(e_{[k_2]})$ and $\hat{f}(e_{[k_1+1]}) < \hat{f}(e_{[k_2+1]})$, then $k_1 \coloneqq k_1 + 1$ and $k_2 \coloneqq k_2 + 1$;
if $\hat{f}(e_{[k_1]}) > \hat{f}(e_{[k_2]})$ and $\hat{f}(e_{[k_1+1]}) > \hat{f}(e_{[k_2+1]})$, then $k_1 \coloneqq k_1 - 1$ and $k_2 \coloneqq k_2 - 1$.

Step 2:
iterate Step 1 until none of the aforementioned conditions is satisfied, or $(k_1 - 1)(k_2 - n) = 0$.

Then, the resulting prediction interval is defined as: $\langle y_{Tforec} + e_{[k_1]},\ y_{Tforec} + e_{[k_2]} \rangle$, where y_{Tp} denotes the direct point forecast for moment T from modal regression, i.e. the subscript $forec$ stands for 'prediction'.

It may be worth mentioning that, under the assumption of unimodality of the estimated density, the above method of constructing prediction intervals can be regarded as a specific approach to finding conditional highest density regions, but here with the help of residuals from a modal regression (for an alternative approach of constructing such regions, see Hyndman, 1995). In our empirical study, we will also use it in combination with other estimation methods and other point forecasts. In kernel density estimation, we will rely on the Gaussian kernel with bandwidth found as in Botev et al. (2010; see also Yao & Li, 2014).

The forecast evaluation in this paper relies on the following measures, computed for y_t being the realizations and y_{tforec} being the forecasts ($t = 1, \dots, T$, where T is the number of forecasts found for a constant forecast horizon):

- the symmetric Mean Absolute Percentage Error (e.g. Makridakis, 1993):

$$sMAPE = \frac{1}{T}\sum_{t}\frac{|y_t - y_{tforec}|}{0.5(y_t + y_{tforec})}; \tag{3}$$

- the modal loss, i.e. the loss function used in estimation (Kemp & Santos Silva, 2012):

$$ModL = 1 - \gamma\frac{1}{T}\sum_{t}\Phi\left(\frac{y_t - y_{tforec}}{h}\right), \tag{4}$$

where Φ is the standard normal density function, $\gamma = \Phi^{-1}(0)$, and h is the bandwidth parameter;

- the set of credibility indexes based on the uniform (rectangular) kernel, defined as:

$$C(b) = \frac{1}{T}\sum_{t}\mathbf{1}\{|y_t - y_{tforec}| < b\}, \tag{5}$$

or, alternatively, as:

$$C(b\%) = \frac{1}{T}\sum_{t}\mathbf{1}\{|y_t - y_{tforec}| < b\%\bar{y}\}, \tag{6}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function, $\bar{y}$ is the arithmetic mean of the data in the estimation sample, and $b > 0$ is a constant.

These measures will indicate forecasting procedures that most often produce forecasts with ex-post forecast errors smaller in magnitude than a certain percent of the mean value or, alternatively, than a certain positive constant.

The creditability measures (5)–(6) can also be defined with other kernel functions that are nonnegative, bounded, and symmetric around zero. Also, please note that the sMAPE is a standard accuracy measure for evaluating mean value forecasts, which was used, for example, in the M3 and M4 forecast competitions (Makridakis et al., 2020; Makridakis & Hibon, 2000). The other measures, on the other hand, are dedicated to measuring the accuracy of modal forecasts.

Besides the evaluation with accuracy measures, we will also perform a battery of statistical tests of forecast quality. They include the test of rationality of modal forecasts by Dimitriadis et al. (2024), the test of equal predictive ability according to Giacomini

and White (2006), and the test of correct unconditional coverage (e.g. Clements, 2005). To operationalise the first two tests, we use the t statistics which have standard normal distributions under the assumption the null hypotheses are true, whereas the third test uses the likelihood ratio statistic, converging under the null hypothesis to the χ_1^2 distribution.

The test of modal rationality by Dimitriadis et al. (2024) is based on the asymptotic identification function for the (conditional) mode. It is assumed that the forecast fulfills certain mixing properties, and that $hT \to \infty$ and $h^7 T \to 0$, where the latter assumptions are standard conditions in nonparametric estimation. Then, under the null hypothesis of an (unconditional) modal rationality of forecasts, the following asymptotic result is obtained:

$$\frac{1}{\sqrt{hT}} \sum_t \Phi'\left(\frac{y_t - y_{tforec}}{h}\right) \Rightarrow N(0, \Omega_{Mode}),$$

where the asymptotic variance is estimated in the following way:

$$\widehat{\Omega}_{Mode} = \frac{1}{hT} \sum_t \left[\Phi'\left(\frac{y_t - y_{tforec}}{h}\right)\right]^2.$$

Later in the empirical part, we perform this test as a final-sample predictive ability test (Clark & McCracken, 2013) and set the bandwidth parameter for rationality testing to the value proposed in Dimitriadis et al. (2024), formula (S.3.1), i.e. $h := k_1 k_2 T^{-\frac{1}{7}}$, where $k_1 = 2.4 Med\left(\left|y_t - y_{tforec} - Med\left(y_t - y_{tforec}\right)\right|\right)$ and $k_2 = e^{-3|\hat{\gamma}|}$ for $\hat{\gamma} = \frac{3\left[\frac{1}{T}\sum_{t=1}^{T}(y_t - y_{tforec}) - Med(y_t - y_{tforec})\right]}{Std(y_t - y_{tforec})}$, with $Med(\cdot)$ and $Std(\cdot)$ denoting the median and the standard deviation.

The Giacomini and White (2006) test is one of the approaches to comparing the predictive properties of competing forecasting solutions based on the average value of the forecast loss differential, i.e. $d_t = L(e_{1t}) - L(e_{2t})$. This test is different from the more popular tests studied in West (2006), because it does not require the estimation error to vanish asymptotically and assumes a fixed estimation window. For this reason, it allows 'the forecasts to be produced by general parametric, semiparametric, and nonparametric estimation techniques' (Giacomini & White, 2006). For further discussion on this test, see Elliot and Timmermann (2016), and the recent polemic in Zhu and Timmermann (2020). As mentioned earlier, we operationalise it here with the t statistic:

$$t = \frac{\bar{d}}{\sqrt{Var(\bar{d})}}, \tag{7}$$

in which the variance in the denominator is estimated using the Newey-West approach (Newey & West, 1987) with the Bartlett kernel bandwidth expanding with the forecast horizon from four (for one-step ahead forecasts) to seven (for four-step ahead forecasts).

Finally, the test of correct unconditional coverage is based on the following test statistic (e.g. Clements, 2005):

$$LR = 2\left\{\ln\left[\left(\frac{N}{T}\right)^{N}\left(1-\frac{N}{T}\right)^{T-N}\right] - \ln[\rho^{N}(1-\rho)^{T-N}]\right\}, \tag{8}$$

where ρ is the nominal coverage, i.e. the assumed confidence level, and the N/T ratio is the sample proportion of hits. As was mentioned earlier, this test uses quantiles of the χ_1^2 distribution as asymptotic critical values.

4. Empirical study

4.1. Data, models, and forecasting procedures

This study focuses on chain indexes of industrial production at constant prices. We use seasonally-adjusted monthly indexes showing the dynamics of production with respect to the previous period. They span the period from January 1994 to December 2019, and concern economies of the following 27 countries: Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Japan, South Korea, Luxembourg, the Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Turkey, the UK and the US. The data were taken from the OECD statistical portal.[3]

Table 1. Descriptive statistics for the indexes of industrial production of the 27 studied economies

(Standardised) moments	Minimum	Maximum	Median	Mean
Mean	99.997	100.744	100.158	100.211
Standard deviation	0.641	6.482	2.147	2.258
Skewness	–2.745	1.253	–0.174	–0.231
Kurtosis	3.303	22.060	5.748	6.387

Source: author's computations.

[3] https://www.oecd.org.

The descriptive statistics in Table 1, jointly with a more detailed examination, demonstrate that the examined production indexes are leptokurtic and often negatively skewed. As noted by Stockhammar and Öller (2011, p. 2038), 'any linear model applied on skewed and leptokurtic data will produce skewed and leptokurtic residuals'. Thus, at least in the case of our linearly-specified models, we can assume that some of the error distributions in our models will not fulfill the Gaussian assumption.

For each of the examined economies, we found point and interval forecasts from one to four steps ahead in a rolling forecasting schema resulting in 100 forecasts for each forecast horizon. In every case, the estimation window was of the size of 209 months.

The following forecasting procedures were employed:

1. The AR(p) model.

   This is the usual autoregressive model for the conditional mean estimated by least squares, with the autoregression order chosen with the AICC, assuming that the minimal and maximal model orders are zero and six. In constructing interval forecasts, we used two approaches (further referred to as AR and AR-Asymmetric). In the former, we assumed that we have a Gaussian autoregressive process, and we constructed prediction intervals on the basis of this assumption and the appropriate formulas for the forecast error variance (constructed assuming the lack of estimation errors). In the latter, we computed the asymmetric prediction intervals presented in Section 3. To this end, we performed the following steps: first, we used a nonparametric bootstrap to generate predictive distributions and then we applied the method presented in Section 3 to deviations of these distributions from the corresponding mean value forecasts. The number of bootstrap samples for each forecast horizon was set at 500. They were generated through random draws from the set of residuals and simulating conditional realizations of the process from one to four periods ahead without accounting for estimation errors. As elsewhere in this paper, the density of forecast errors was estimated with the nonparametric kernel method introduced by Botev et al. (2010).

2. The modal autoregression model (denoted as MAR).

   The modal autoregression model of order p, denoted MAR(p), is defined in the following way:

$$Mode(y_t|\Omega_{t-1}) = \beta_0^j + \beta_1^j y_{t-1} + \ldots + \beta_p^j y_{t-p} \tag{9}$$

   and is estimated using the modal regression estimator as described in Section 2. It is used to compute one-step-ahead modal forecasts. To allow the computation of multiple-step-ahead modal forecasts, we also considered a more general linear specification in the modal regression framework, namely the following model:

$$Mode(y_{t+j}|\Omega_{t-1}) = \beta_0^j + \beta_1^j y_{t-1} + \beta_2^j y_{t-2} \tag{10}$$

with $j > 0$, which made it possible to find direct modal forecasts from two to four steps ahead. This model was estimated separately for each forecast horizon, except for $p = 0$, in which case the same forecasts were assumed at each horizon. The autoregression order was set to that of the standard autoregressive model, and the same set of regressors was used for all $j \geq 0$. Interval forecasts were found using the method of Yao and Li (2014), as described in the previous section. In short: after finding point forecasts up to four steps ahead, residuals from each of the estimated regressions were computed and applied directly to find the asymmetric prediction intervals.

3. The AR(p) model with nonparametric modal correction (denoted as AR-Mode).

   This is the same autoregressive model for the conditional mean as that in the first approach, but combined in the second step with estimating the mode of predictive distribution. As before, 500 simulated conditional realizations of the process for each forecast horizon were used. The mode was estimated through a modal regression approach assuming that the set of regressors consists of a constant only. In interval forecasting, the mode was subtracted from the generated predictive distribution, and the differences were subsequently inserted in the algorithm of finding asymmetric prediction intervals presented in Section 3.

4. Unconditional modal forecasts (denoted as Mode).

   The computations here are similar to AR-Mode, but the order of autoregression is set to zero. The same point and interval forecasts are used for each horizon.

5. The quantile autoregression model (denoted as QAR).[4]

   This is an autoregressive model for the conditional quantile of order 0.5, estimated with the quantile regression estimator separately for each forecast horizon (similar to the MAR model, except if $p = 0$), assuming that the set of regressors is the same for all horizons. The autoregression order was the same as in the case of the autoregressive model for the mean. To find interval forecasts, residuals from estimation were used in the algorithm presented in Section 3, in a similar way to what was done for modal autoregression.

6. The Momentum-Threshold Autoregressive (M-TAR) model for the conditional mean.

   This is the Momentum-TAR model, introduced by Enders and Granger (1998), but with the long difference $\Delta_2 y_{t-1} = y_{t-1} - y_{t-3} = \Delta y_{t-1} + \Delta y_{t-2}$ as the threshold

[4] For a rigorous presentation of the concepts of quantile regression and autoregression, see Koenker and Bassett (1978) and Koenker and Xiao (2006). Please note, however, that in this paper, in the computations of multiple-step-ahead forecasts, we do not assume that parameters of the linear autoregression model are changing across quantiles.

variable. The threshold value was set to zero, and the autoregression order in each regime was the same and equal to that of the linear autoregressive model, i.e., we estimated the following model by least squares:

$$E(y_t|\Omega_{t-1}) = \begin{cases} \beta_{10} + \beta_{11}y_{t-1} + \ldots + \beta_{1p}y_{t-p} & \text{for } \Delta_2 y_{t-1} \geq 0, \\ \beta_{20} + \beta_{21}y_{t-1} + \ldots + \beta_{2p}y_{t-p} & \text{for } \Delta_2 y_{t-1} < 0. \end{cases} \tag{11}$$

The model (11) assumes a different dynamic of growth depending on whether the most recent data show a local upward or downward tendency. Thus, it enables modelling asymmetries over the business cycle consisting of alternating phases of expansions and recessions. The long difference in the definition of the threshold variable was used to make regime switching less sensitive to the noise. Point forecasts of the mean value were obtained through the nonparametric bootstrap with 500 simulated conditional realizations. Prediction intervals were constructed using the method presented in Section 3.

7. The modal M-TAR model (denoted as MM-TAR).

This is the M-TAR model as defined above, but estimated as a modal regression model separately for each horizon, i.e.:

$$Mode(y_{t+j}|\Omega_{t-1}) = \begin{cases} \beta_{10}^j + \beta_{11}^j y_{t-1} + \ldots + \beta_{1p}^j y_{t-p} & \text{for } \Delta_2 y_{t-1} \geq 0, \\ \beta_{20}^j + \beta_{21}^j y_{t-1} + \ldots + \beta_{2p}^j y_{t-p} & \text{for } \Delta_2 y_{t-1} < 0, \end{cases} \tag{12}$$

where $j = 0, 1, 2, 3$.

As previously, the autoregression order was set to that of the standard autoregressive model. Prediction intervals were found in the same way as in the case of linear modal autoregression.

8. The AR(p)-GARCH(1, 1) model (denoted as AR-GARCH).

This is the usual autoregressive model with a GARCH(1, 1) equation for the conditional variance and Gaussian innovations, i.e.:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \varepsilon_t, \\ \varepsilon_t &= \sqrt{h_t}\nu_t, \quad \nu_t \sim N(0,1), \\ h_t &= \omega + \gamma h_{t-1} + \alpha\varepsilon_{t-1}^2. \end{aligned} \tag{13}$$

This model was estimated using the Econometrics Toolbox for Matlab assuming the interior-point algorithm. The order of autoregression was the same as in the case

of linear autoregression without GARCH errors. Mean value forecasts were computed, which, in the case of multiple-step-ahead predictions, turned out to be using nonparametric bootstrap and 500 conditional realizations of the AR(p)-GARCH(1, 1) process. This means that the Gaussian assumption was used exclusively for estimation. Interval forecasts were also generated using the simulation approach, combined again with the method of constructing asymmetric prediction intervals.

9. The M-TAR(p, p)-GARCH(1, 1) model (denoted as M-TAR-GARCH).

In this model, the conditional mean equation is specified in the same way as in the case of the M-TAR model, while the other components are the same as in AR-GARCH, i.e.:

$$\begin{gathered} y_t = \begin{cases} \beta_{10} + \beta_{11} y_{t-1} + \ldots + \beta_{1p} y_{t-p} + \varepsilon_t & \text{for } \Delta_2 y_{t-1} \geq 0, \\ \beta_{20} + \beta_{21} y_{t-1} + \ldots + \beta_{2p} y_{t-p} + \varepsilon_t & \text{for } \Delta_2 y_{t-1} < 0, \end{cases} \\ \varepsilon_t = \sqrt{h_t} \nu_t, \quad h_t = \omega + \gamma h_{t-1} + \alpha \varepsilon_{t-1}^2, \quad \nu_t \sim N(0,1). \end{gathered} \tag{14}$$

As previously, the estimation was based on the interior-point algorithm, while bootstrap was used to generate 500 conditional realizations of the M-TAR(p, p)-GARCH(1, 1) process without referring to the Gaussian assumption. Also, prediction intervals were found using the same approach as in the previous case.

4.2. Point forecasts

Our assessment of point forecasts in terms of accuracy is based on the sMAPE and the credibility measure (5) computed assuming different values of parameter b. The ratios of mean values of $C(b)$ reported in Table 2 are for b equal 3, 2.5, 2, and 1.5.[5] These measures show which of the forecasting procedures most often produced forecasts of growth with ex post forecast errors smaller than ±3%, 2.5%, 2%, and 1.5%.[6] Furthermore, Table 3 presents information about the results of tests of equal predictive ability corresponding with the measures in Table 2. The level of significance was set to 5%. Table 3 also summarizes the outcomes of the two-sided tests of modal rationality of forecasts, also performed at the 5-percent level of significance.

[5] The accuracy measures shown in Table 2 are in robustified versions. In the case of the sMAPEs, which are defined here as averages across time and space, a double trimming (first in the time and then in the spatial dimension) was applied, cutting off 5% of the smallest and 5% of the largest values in each dimension. In the case of $C(b)$,the trimming was in the spatial dimension, again cutting off 5% of the lowest and highest values.

[6] These values of b resulted in average empirical credibility indexes for different forecasting procedures and forecast horizons oscillating around 90%, 83%, 75%, and 63%, respectively. For $b \leq 1$, the average values of these measures were almost exclusively below 50%.

Table 2. Ratios of robustified symmetric Mean Absolute Percentage Errors and credibility indexes for point forecasts of industrial output indexes; aggregated outcomes for the 27 studied economies

Forecast horizon	AR	AR-Mode	Mode	QAR	M-TAR	MM-TAR	AR-GARCH	M-TAR-GARCH
Ratios of doubly trimmed sMAPEs								
1	**0.991**	**0.985**	**0.918**	**0.994**	**0.988**	**0.997**	**0.996**	**0.992**
2	1.007	1.001	1.006	**0.998**	**0.997**	**0.986**	**1.000**	**0.994**
3	1.006	**0.996**	1.004	1.004	**0.990**	**0.993**	1.002	1.002
4	1.009	1.005	1.005	1.002	**0.991**	**0.984**	1.006	1.006
Ratios of robustified means of *C*(3)								
1	**1.000**	**1.003**	**1.021**	**1.000**	**1.002**	**1.000**	0.997	**1.000**
2	**1.001**	**1.000**	0.995	**1.001**	**1.004**	**1.007**	**1.003**	**1.002**
3	0.998	**1.002**	0.998	**1.001**	**1.001**	**1.002**	**1.000**	0.999
4	0.996	**1.003**	**1.000**	**1.001**	**1.000**	**1.001**	0.999	**1.003**
Ratios of robustified means of *C*(2.5)								
1	**1.004**	**1.008**	**1.041**	**1.002**	**1.008**	**1.007**	0.999	**1.002**
2	**1.001**	**1.004**	0.996	**1.001**	**1.008**	**1.010**	**1.000**	**1.005**
3	**1.006**	**1.001**	0.998	**1.003**	**1.011**	**1.002**	**1.006**	**1.004**
4	**1.000**	0.998	0.998	0.998	**1.006**	**1.003**	**1.000**	**1.002**
Ratios of robustified means of *C*(2)								
1	**1.002**	**1.003**	**1.035**	0.999	**1.020**	**1.009**	0.996	**1.012**
2	0.999	**1.006**	0.999	**1.000**	**1.001**	**1.019**	**1.003**	**1.013**
3	0.994	**1.009**	0.997	0.996	**1.008**	**1.011**	0.994	**1.004**
4	0.995	0.999	0.999	**1.004**	**1.004**	**1.012**	0.997	**1.001**
Ratios of robustified means of *C*(1.5)								
1	**1.009**	**1.014**	**1.041**	**1.007**	**1.005**	**1.001**	**1.007**	0.996
2	0.994	0.998	0.999	**1.010**	**1.010**	**1.025**	0.999	**1.006**
3	0.995	0.993	0.992	0.996	**1.005**	**1.008**	0.997	0.999
4	0.990	0.990	0.997	0.998	**1.019**	**1.008**	**1.001**	0.996

Note. Benchmark is the modal autoregression (MAR) with the same order of autoregression as in the case of AR, separate (direct) estimation for each forecast horizon; AR denotes the AR(p) model under the least square estimation; AR-Mode means the AR(p) model estimated in the same way as in the case of AR, but with forecasts computed as the nonparametrically estimated mode of the predictive distributions; Mode is the same forecasts for each horizon equal to the nonparametrically estimated mode; QAR denotes quantile autoregression for the quantile of order 0.5 with the same order of autoregression as in the AR, a separate (direct) estimation for each forecast horizon; M-TAR stands for the M-TAR model under least squares estimation with the threshold variable being the long difference and the threshold value set to 0, the order of autoregression as in AR; MM-TAR signifies the modal M-TAR model specified as in the case of M-TAR, separate (direct) estimation for each forecast horizon; AR-GARCH is the same as AR but with added GARCH(1, 1) equation with Gaussian innovations; M-TAR-GARCH is the same as M-TAR but with added GARCH(1, 1) equation with Gaussian innovations. The presented numbers are the ratios of the appropriate accuracy measures for the benchmark modal autoregression model to the corresponding measures for the eight alternative methods. All instances where the benchmark modal autoregression performed no worse than an alternative approach are given in bold. Please refer to the text for more information.

Source: author's computations.

When analysing the relative aggregate accuracy measures collated in Table 2, we first notice that, somewhat unexpectedly, linear modal autoregression performs best

of all the forecasting procedures in terms of the sMAPE when forecasting one step ahead, and the modal M-TAR model is the second best (see the first panel of Table 2). This results from the robustness properties of modal regression. Please note that in Kemp et al. (2020), an MSE-based forecast loss function for a three-variable vector autoregression for inflation, unemployment rate, and interest rate is also minimised when a multivariate extension of modal estimation is applied. At all the other horizons, however, MAR is outperformed by the standard AR model, which generates the best forecasts two, three, and four steps ahead of all the examined methods. This is contrary to the empirical outcomes in Kemp et al. (2020), where the best multiple-step-ahead forecasts were obtained by means of the Vector Mode Regression. Interestingly, however, the procedure denoted as Mode, which is based on MAR models of order zero, in our study also generates very good forecasting outcomes when forecasting more than one step ahead.

To a large extent, the aggregate credibility indexes in Table 2 also favour the MAR model in forecasting one step ahead. In particular, MAR performs best of all the linearly-specified models (AR, AR-Mode, Mode, and QAR) in terms of $C(2.5)$ and $C(1.5)$ (see the middle and bottom panels of Table 2), although the differences in their credibility are rather small. However, our experiment with smaller values of b shows that the credibility indexes of MAR can then be much better than those of AR and QAR. In particular, for b set to 0.25, the ratios of robustified means of credibility measures are equal to 1.032 for both AR and QAR.

Similarly to the assessment by means of sMAPE, we observe a deterioration in the relative credibility of linear modal autoregression as the forecast horizon becomes longer, although this phenomenon occurs to a lesser extent than in the previous case, at least as regards $C(3)$ and $C(2.5)$. Furthermore, when considering all forecasting horizons, AR-GARCH achieves, on average, a similar credibility to MAR, while AR outperforms it in terms of $C(2)$ and $C(1.5)$ when forecasting from two to four steps ahead. On the other hand, however, the MAR of order zero (i.e. the method denoted as Mode) produces, on average, the best multiple-step-ahead forecasts as regards $C(3)$ and $C(2.5)$ (see the corresponding rows in Table 2).

Besides, with a few noticeable exceptions (in particular, $C(1.5)$ when forecasting one and four steps ahead with M-TAR-GARCH), the threshold models seem not to improve the trustworthiness of macroeconomic forecasts beyond that of the simple MAR model. In addition, the standard AR model with the modal correction (AR-Mode) is most often outperformed by the same model without the nonparametric adjustment.[7]

[7] For this reason, we decided not to apply the modal correction to nonlinear (in mean and/or variance) specifications.

Thus, we conclude that, in terms of credibility (trustworthiness), linear modal autoregression often successfully competes with other linearly-specified models, but achieves a similar accuracy to linear autoregression with a GARCH equation, unless the autoregression order is set to zero when forecasting multiple steps ahead. On the other hand, however, it should be emphasised that sMAPE, currently a highly popular forecast accuracy measure (especially in forecast competitions), indicates direct MAR models as the best approach to forecasting output indexes one step ahead.

Table 3. Results of tests of the quality of industrial output forecasts; aggregated outcomes for the 27 studied economies

Forecast horizon	MAR	AR	AR-Mode	Mode	QAR	M-TAR	MM-TAR	AR-GARCH	M-TAR-GARCH
Equal predictive ability: credibility index *C*(3)									
1	.	0/0	1/0	3/0	0/0	0/0	0/0	1/1	0/0
2	.	0/0	1/1	0/1	0/0	1/0	2/0	1/0	1/0
3	.	0/0	0/0	0/0	0/0	1/1	1//0	1/0	1/0
4	.	0/2	3/2	1/3	1/0	1/0	2/0	2/1	3/0
Equal predictive ability: credibility index *C*(2.5)									
1	.	3/1	4/1	8/0	1/0	2/1	1/0	1/0	0/0
2	.	0/0	1/0	1/0	1/0	0/0	0/1	1/0	1/0
3	.	2/0	1/0	0/1	0/0	3/0	1/0	2/0	0/0
4	.	0/1	0/2	0/0	0/1	3/2	2/0	0/1	0/1
Equal predictive ability: credibility index *C*(2)									
1	.	1/2	1/1	4/0	0/0	5/0	1/1	2/2	3/0
2	.	1/1	1/1	2/1	2/1	2/1	6/0	2/0	3/0
3	.	0/3	0/1	1/1	0/1	1/0	3/0	1/4	1/2
4	.	1/0	2/1	1/0	4/1	2/2	5/0	1/2	0/0
Equal predictive ability: credibility index *C*(1.5)									
1	.	3/1	3/1	2/1	2/0	2/3	1/1	1/1	0/1
2	.	1/2	2/2	2/0	3/0	2/3	4/1	3/0	5/2
3	.	0/2	3/4	0/1	2/1	2/1	1/0	1/1	0/1
4	.	1/4	0/2	0/2	1/1	3/2	4/2	2/3	1/1
Rationality of modal forecasts									
1	0	1	1	0	2	1	1	0	0
2	2	1	1	0	3	0	0	0	0
3	0	2	0	0	1	1	0	0	0
4	0	1	0	0	0	1	0	0	0
Equal predictive ability (sMAPE)									
1	.	2/3	5/2	16/1	4/1	5/1	5/1	1/1	1/0
2	.	0/2	4/1	2/4	0/5	3/2	2/0	1/2	2/0
3	.	0/1	2/0	3/2	3/3	3/1	1/1	1/1	2/1
4	.	0/5	2/3	2/4	2/3	2/1	4/0	1/6	0/5

Note. Equal predictive ability (credibility) – the presented numbers are the amounts of instances where the credibility of the benchmark modal autoregression model is significantly above or below the credibility of the alternative approaches, the level of significance at 5%; forecast rationality tests – the presented numbers are amounts the rejections of modal rationality among the examined time series at the significance level of 5% in two-sided tests; equal predictive ability (sMAPE) – the presented numbers are the amounts instances where the sMAPE of the benchmark modal autoregression model is significantly below or above the sMAPE of the alternative approaches, the 5-percent level of significance. See note to Table 2 and the text for more details.

Source: author's computations.

Moreover, when examining the results of statistical tests in Table 3, at first we notice that all the forecasting procedures produce forecasts which can be rationalised as modal forecasts (see the fifth panel of Table 3). But even more importantly, we can also observe that, with a few noticeable exceptions (in particular, sMAPE and $C(2.5)$ when forecasting one step ahead with the 'Mode' method), the tests of equal predictive ability do not visibly favour linear modal autoregression.

4.3. Interval forecasts

Our assessment of interval forecasts is performed with tests of correct unconditional coverage, mean coverage statistics, and mean ratios of length of prediction intervals.[8] The 5-percent level of significance is assumed in statistical verifications. The obtained results for the nominal confidence levels of 90%, 80%, and 50% are presented in Table 4, while further statistics (for the nominal confidence levels of 95%, 70%, and 60%) are briefly discussed in the text.[9]

It is worth observing that the mean ratios of length indicate that linear-modal-autoregression estimation can indeed lead at all the examined horizons to shorter prediction intervals than other forecasting procedures applied to linearly specified models (and also the nonlinear autoregressive model for the conditional mean), but this is more likely for the lower nominal confidence levels of 80%, 70%, 60%, and 50%, and may not hold true for high confidence levels (see the bottom panels of Table 4). In particular, the 95-percent prediction intervals obtained with linear modal autoregression are in mean terms shorter only than the intervals obtained with the Mode approach and, in one-step-ahead forecasting, also as compared with the intervals generated through quantile autoregression (QAR; not reported).

Except for the intervals constructed under the Gaussian assumption, the difference in length of the prediction intervals is not substantial. The observed mean ratios of length are all below 1.04, which means that linearly-specified models other than MAR and the M-TAR lead to interval forecasts with the lengths on average larger than those of the MAR-based intervals by less than 4% (see the bottom panels of Table 4).

On the other hand, if we also include the remaining three models (MM-TAR, AR-GARCH, and M-TAR-GARCH) in our comparison, we will see that the models with GARCH errors largely outperform all the other strategies. This is seen across all the statistics collated in Table 4. Namely, the models with GARCH equations most often produce the shortest prediction intervals in mean terms, which also have very good coverage properties. The latter is confirmed by both the mean coverage statistics, which are close to the nominal coverage levels, and by the results of the tests of the correct unconditional coverage, which rarely reject the null of correct coverage in this case.

---

[8] Mean ratios are given in robustified versions obtained through a double trimming performed in a way similar to the robustified sMAPE used to compute the ratios presented in Table 2.

[9] The test based on the statistics (8) could not be performed for the 95-percent prediction intervals, because the number of hits (N) was in this case often equal to the number of forecasts (T).

Table 4. Properties of interval forecasts of industrial output; aggregated outcomes for the 27 studied economies

Forecast horizon	MAR	AR	AR-Asymmetric	AR-Mode	Mode	QAR	M-TAR	MM-TAR	AR-GARCH	M-TAR-GARCH
Mean coverage, 1 – α = 90%										
1	0.926	0.932	0.919	0.918	0.930	0.923	0.911	0.917	**0.902**	**0.898**
2	0.923	0.931	0.914	0.903	0.930	0.923	0.910	0.915	0.903	**0.901**
3	0.922	0.933	0.920	0.920	0.931	0.922	0.913	0.915	0.903	**0.901**
4	0.921	0.933	0.920	0.917	0.928	0.924	**0.909**	0.916	0.917	0.916
Mean coverage, 1 – α = 80%										
1	0.828	0.869	0.824	0.825	0.835	0.825	0.820	0.812	0.809	**0.798**
2	0.817	0.866	0.823	**0.800**	0.839	0.819	0.817	0.811	0.802	0.802
3	0.821	0.876	0.828	0.819	0.838	0.823	0.825	0.816	**0.810**	0.807
4	0.823	0.877	0.833	0.820	0.835	0.826	0.820	**0.813**	0.816	0.817
Mean coverage, 1 – α = 50%										
1	0.514	0.600	0.519	0.513	0.519	0.509	0.522	0.510	0.506	**0.503**
2	0.514	0.602	0.520	0.490	0.522	0.517	0.524	0.494	0.494	**0.499**
3	0.516	0.599	0.529	0.507	0.524	0.516	0.524	0.505	**0.499**	0.504
4	0.509	0.604	0.526	0.512	0.523	0.514	0.511	**0.494**	0.507	0.508
Unconditional coverage, 1 – α = 90%										
1	10	11	8	9	9	9	8	7	3	4
2	7	11	7	8	10	10	8	9	3	3
3	12	13	7	7	11	9	8	9	6	4
4	8	14	11	6	11	8	7	8	4	5
Unconditional coverage, 1 – α = 80%										
1	6	12	6	7	8	6	9	7	1	2
2	8	13	7	5	9	8	6	6	3	2
3	7	14	6	8	9	7	8	7	4	2
4	5	14	8	5	7	6	6	8	2	1
Unconditional coverage, 1 – α = 50%										
1	7	15	6	6	4	5	7	4	4	4
2	5	16	8	4	3	4	8	4	2	5
3	5	17	7	5	3	6	5	4	2	3
4	3	14	4	3	4	4	6	3	2	3
Robustified mean ratios of length, 1 – α = 90%										
1	.	**1.062**	**1.000**	**1.000**	**1.083**	**1.000**	0.994	0.992	0.946	0.934
2	.	**1.048**	0.981	0.980	**1.020**	**1.001**	0.980	0.983	0.942	0.943
3	.	**1.050**	0.984	0.984	**1.015**	**1.000**	0.983	0.990	0.955	0.956
4	.	**1.051**	0.987	0.986	**1.010**	**1.003**	0.980	0.981	0.963	0.964
Robustified mean ratios of length, 1 – α = 80%										
1	.	**1.128**	**1.011**	**1.011**	**1.080**	**1.001**	**1.008**	0.989	0.967	0.952
2	.	**1.121**	**1.005**	**1.005**	**1.023**	**1.005**	**1.011**	0.989	0.957	0.963
3	.	**1.124**	**1.008**	**1.008**	**1.020**	**1.000**	**1.016**	0.986	0.969	0.974
4	.	**1.130**	**1.018**	**1.016**	**1.018**	**1.001**	**1.017**	0.991	0.979	0.985
Robustified mean ratios of length, 1 – α = 50%										
1	.	**1.219**	**1.022**	**1.022**	**1.088**	**1.003**	**1.026**	0.984	0.987	0.986
2	.	**1.198**	**1.014**	**1.014**	**1.016**	**1.004**	**1.027**	0.981	0.965	0.968
3	.	**1.209**	**1.025**	**1.025**	**1.018**	**1.002**	**1.039**	0.977	0.977	0.983
4	.	**1.212**	**1.033**	**1.032**	**1.014**	**1.005**	**1.038**	0.985	0.984	0.991

Note. Mean coverage – the values which are closest to the nominal coverage of 90%, 80%, and 50% are bolded; mean ratios of length – the mean values (for forecast periods and countries) of the ratios of lengths of the prediction intervals obtained with one of the eight alternative methods to those obtained with the benchmark modal autoregression model, the instances when modal autoregression produced prediction intervals shorter in mean terms are bolded; unconditional coverage – the presented numbers are the amounts of rejections of correct coverage for the 90%, 80%, and 50% prediction intervals at the significance level of 5%. See note to Table 2 and the text for more details.

Source: author's computations.

There is, however, an interesting exception, namely the 80-percent, 60-percent and 50-percent prediction intervals at horizon four, where both models with GARCH errors are outperformed by nonlinear modal autoregression (MM-TAR) as regards the mean coverage statistics. In the case of the confidence levels of 60% and 50%, this model also competes successfully with AR-GARCH at horizon one in terms of the mean ratios of length. Generally, as regards the confidence levels of 80% and below and the mean ratios of length and mean coverage statistics, MM-TAR is the third best model, preceded by two with GARCH errors. Interestingly, for the 90-percent and 95-percent intervals, a similar outcome is achieved with the M-TAR model for the mean value, which is the third best in terms of mean coverage at horizon four. Thus, assuming nonlinearity both in mean and in mode can help to produce better interval forecasts, even without specifying a GARCH equation. Somewhat surprisingly (in the empirical context), these types of nonlinearity appear to be much more important in computing interval forecasts than in computing point forecasts. In particular, when looking at the mean coverage statistics at the longest horizon considered here and except for 70-percent intervals, either MM-TAR or M-TAR outperforms all the other solutions.

It is also worth noticing that all the examined forecasting strategies tend towards overcoverage. There are only a few exceptions to this, namely when forecasting up to two or three steps ahead with one or both of the models with GARCH equations, and forecasting with MM-TAR at the confidence levels of 60% and 50%. Moreover, the M-TAR model with GARCH errors provides the shortest interval forecasts with satisfactory coverage at the 95-percent confidence level, clearly outperforming linear and nonlinear modal autoregression, especially in terms of the lengths of prediction intervals (not reported).

5. Conclusions

Modal regression is a promising statistical approach, having potentially much to offer also in time series analysis and forecasting. In this paper, we draw attention to the forecasting benefits of this approach in the macroeconomic context. Namely, we have undertaken the task of checking if simple linear or nonlinear modal autoregression could help produce macroeconomic forecasts with certain desirable properties, jointly labelled here as 'credibility' or, alternatively, 'trustworthiness'. To be more precise, we have asked if modal (auto)regression could be employed to generate forecasts with low values of bounded loss functions symmetric around zero, in this study mainly represented by the rectangular loss function, and to obtain prediction intervals narrower than by means of other approaches. To the best of our knowledge, our empirical study, based on industrial output indexes in a large panel of OECD

countries, has been the first attempt to provide answers to such questions in the economic context.

This study demonstrates that modal autoregression, thanks to its robustness properties, leads to more accurate one-step-ahead point forecasts of industrial output in terms of the sMAPE than those yielded by other approaches. These forecasts also result in high values of credibility indexes, although the credibility and loss differentials are rarely statistically significant. In addition, modal estimation is more likely to provide narrow prediction intervals in the case of lower nominal confidence levels such as 80% and 50%. Even then, however, it rarely outperforms the standard (linear or nonlinear) autoregressive models with GARCH errors, which provide the narrowest (and, at the same time, good-quality) interval forecasts, although with a few noticeable exceptions. On the other hand, however, the linear modal autoregression model, on which this study mostly focuses, outperforms other linearly-specified models for measures of central tendency, i.e. median and mean autoregression, when considering confidence levels of 80% and below. This can partly result from the nonlinearity of output indexes, including their regime-switching dynamics and heteroscedasticity features.

Thus, modal estimation has much to offer both in terms of minimising simple forecast loss functions and generating short prediction intervals. It must be noted, however, that our observations are limited to the forecasting of economic output and therefore may not apply to a different empirical context, e.g. different output measures and time intervals of an analysis. They can also be influenced by the implementation schemas of modal regression and our definitions of bounded and symmetric loss functions.

Acknowledgements

The paper is part of a research project financed by the National Science Center, Kraków, Poland (grant no. 2017/27/B/HS4/01025). I would like to thank Weixin Yao for sharing with me his Matlab codes used in Yao and Li (2014), which the author adjusted appropriately for this study.

References

Berger, D., Dew-Becker, I., & Giglio, S. (2020). Uncertainty Shocks as Second-Moment News Shocks. *The Review of Economic Studies*, *87*(1), 40–76. https://doi.org/10.1093/restud/rdz010.

Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). Kernel Density Estimation via Diffusion. *The Annals of Statistics*, *38*(5), 2916–2957. https://doi.org/10.1214/10-aos799.

Bruzda, J. (2007). *Procesy nieliniowe i zależności długookresowe w ekonomii*. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.

Bruzda, J. (2019). Quantile Smoothing in Supply Chain and Logistics Forecasting. *International Journal of Production Economics*, *208*, 122–139. https://doi.org/10.1016/j.ijpe.2018.11.015.

Chauvet, M., & Potter, S. (2013). Forecasting Output. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (vol. 2A, pp. 141–194). North-Holland. https://doi.org/10.1016/B978-0-444-53683-9.00003-7.

Chen, Y.-C., Genovese, C. R., Tibshirani, R. J., & Wasserman, L. (2016). Nonparametric Modal Regression. *Annals of Statistics*, *44*(2), 489–514. https://doi.org/10.1214/15-aos1373.

Clark, T., & McCracken, M. (2013). Advances in Forecast Evaluation. In G. Elliot & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (vol. 2B, pp. 1107–1203). North-Holland. https://doi.org/10.1016/B978-0-444-62731-5.00020-8.

Clements, M. P. (2005). *Evaluating Econometric Forecasts of Economic and Financial Variables*. Palgrave Macmillan. https://doi.org/10.1057/9780230596146.

Dimitriadis, T., Patton, A. J., & Schmidt, P. W. (2024). *Testing Forecast Rationality for Measures of Central Tendency*. https://arxiv.org/pdf/1910.12545.

Einbeck, J., & Tutz, G. (2006). Modelling beyond regression functions: an application of multimodal regression to speed-flow data. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, *55*(4), 461–475. https://doi.org/10.1111/j.1467-9876.2006.00547.x.

Elliott, G., & Timmermann, A. (2008). Economic Forecasting. *Journal of Economic Literature*, *46*(1), 3–56. https://doi.org/10.1257/jel.46.1.3.

Elliott, G., & Timmermann, A. (2016). *Economic Forecasting*. Princeton University Press.

Enders, W., & Granger, C. W. J. (1998). Unit-Root Tests and Asymmetric Adjustment With an Example Using the Term Structure of Interest Rates. *Journal of Business & Economic Statistics*, *16*(3), 304–311. https://doi.org/10.2307/1392506.

Fagiolo, G., Napoletano, M., & Roventini, A. (2008). Are Output Growth-Rate Distributions Fat-Tailed? Some Evidence from OECD Countries. *Journal of Applied Econometrics*, *23*(5), 639–669. https://doi.org/10.1002/jae.1003.

Feng, Y., Fan, J., & Suykens, J. A. K. (2020). A Statistical Learning Approach to Modal Regression. *Journal of Machine Learning Research*, *21*(1), 25–59. https://dl.acm.org/doi/10.5555/3455716.3455718.

Gambetta, D. (1988). Can We Trust Trust?. In D. Gambetta (Ed.), *Trust. Making and Breaking Cooperative Relations* (pp. 213–237). Basil Blackwell.

Giacomini, R., & White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica*, *74*(6), 1545–1578. https://doi.org/10.1111/j.1468-0262.2006.00718.x.

Gneiting, T. (2011). Making and Evaluating Point Forecasts. *Journal of the American Statistical Association*, *106*(494), 746–762. https://doi.org/10.1198/jasa.2011.r10138.

Granger, C. W. J., & Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford University Press. https://doi.org/10.1093/oso/9780198773191.001.0001.

Heinrich, C. (2014). The mode functional is not elicitable. *Biometrika*, *101*(1), 245–251. https://doi.org/10.1093/biomet/ast048.

Heinrich-Mertsching, C., & Fissler, T. (2021). Is the Mode Elicitable Relative to Unimodal Distributions?. *Biometrika*, *109*(4), 1157–1164. https://doi.org/10.1093/biomet/asab065.

Hyndman, R. J. (1995). Highest-density Forecast Regions for Non-linear and Non-normal Time Series Models. *Journal of Forecasting*, *14*(5), 431–441. https://doi.org/10.1002/for.3980140503.

Kemp, G. C. R., Parente, P. M. D. C., & Santos Silva, J. M. C. (2020). Dynamic Vector Mode Regression. *Journal of Business and Economic Statistics*, 38. https://doi.org/10.1080/07350015.2018.1562935.

Kemp, G. C. R., & Santos Silva, J. M. C. (2012). Regression Towards the Mode. *Journal of Econometrics*, *170*(1), 92–101. https://doi.org/10.1016/j.jeconom.2012.03.002.

Koenker, R., & Bassett, G. Jr. (1978). Regression Quantiles. *Econometrica*, *46*(1), 33–50. https://doi.org/10.2307/1913643.

Koenker, R., & Xiao, Z. (2006). Quantile Autoregression. *Journal of the American Statistical Association*, *101*(475), 980–990. https://doi.org/10.1198/016214506000000672.

Krief, J. M. (2017). Semi-linear mode regression. *The Econometrics Journal*, *20*(2), 149–167. https://doi.org/10.1111/ectj.12088.

Kröger, S., & Pierrot, T. (2019). *What point of a distribution summarises point predictions?* (WZB Discussion Paper No. SP II 2019-212). http://hdl.handle.net/10419/206533.

Lee, M. J. (1989). Mode Regression. *Journal of Econometrics*, *42*(3), 337–349. https://doi.org/10.1016/0304-4076(89)90057-2.

Makridakis, S. (1993). Accuracy Measures: Theoretical and Practical Concerns. *International Journal of Forecasting*, *9*(4), 527–529. https://doi.org/10.1016/0169-2070(93)90079-3.

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476. https://doi.org/10.1016/S0169-2070(00)00057-1.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014.

Manski, C. F. (1991). Regression. *Journal of Economic Literature*, *29*(1), 34–50.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708. https://doi.org/10.2307/1913610.

Pońsko, P., & Rybarczyk, B. (2016). *Fan Chart – a tool for NBP's monetary policy making* (NBP Working Paper No. 241). https://static.nbp.pl/publikacje/materialy-i-studia/241_en.pdf.

Sager, T. W., & Thisted, R. A. (1982). Maximum Likelihood Estimation of Isotonic Modal Regression. *The Annals of Statistics*, *10*(3), 690–707. https://doi.org/10.1214/aos/1176345865.

Stockhammar, P., & Öller, L.-E. (2011). On the Probability Distribution of Economic Growth. *Journal of Applied Statistics*, *38*(9), 2023–2041. https://doi.org/10.1080/02664763.2010.545110.

Temple, J. (1999). The New Growth Evidence. *Journal of Economic Literature*, *37*(1), 112–156. https://doi.org/10.1257/jel.37.1.112.

Ul Hassan, M., & Stockhammar, P. (2016). Fitting Probability Distributions to Economic Growth: A Maximum Likelihood Approach. *Journal of Applied Statistics*, *43*(9), 1583–1603. https://doi.org/10.1080/02664763.2015.1117586.

Ullah, A., Wang, T., & Yao, W. (2021). Modal Regression for Fixed Effects Panel Data. *Empirical Economics*, *60*(1), 261–308. https://doi.org/10.1007/s00181-020-01999-w.

Ullah, A., Wang, T., & Yao, W. (2022). Nonlinear modal regression for dependent data with application for predicting COVID-19. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(3), 1424–1453. https://doi.org/10.1111/rssa.12849.

Ullah, A., Wang, T., & Yao, W. (2023). Semiparametric Partially Linear Varying Coefficient Modal Regression. *Journal of Econometrics*, *235*(2), 1001–1026. https://doi.org/10.1016/j.jeconom.2022.09.002.

Wang, T. (n.d.). Parametric Modal Regression with Autocorrelated Error Process. *Statistica Sinica*. Advance online publication. https://doi.org/10.5705/ss.202021.0405.

Weiss, A. A. (1996). Estimating Time Series Models Using the Relevant Cost Function. *Journal of Applied Econometrics*, *11*(5), 539–560. https://doi.org/10.1002/(SICI)1099-1255(199609)11:5%3C539::AID-JAE412%3E3.0.CO;2-I.

West, K. D. (2006). Forecast Evaluation. In G. Elliot, C. W. J. Granger & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (vol. 1, 99–134). North-Holland. https://doi.org/10.1016/S1574-0706(05)01003-7.

Xiang, S., & Yao, W. (2022). Nonparametric statistical learning based on modal regression. *Journal of Computational and Applied Mathematics*, *409*, 1–15. https://doi.org/10.1016/j.cam.2022.114130.

Yao, W., & Li, L. (2014). A New Regression Model: Modal Linear Regression. *Scandinavian Journal of Statistics*, *41*(3), 656–671. https://doi.org/10.1111/sjos.12054.

Yao, W., Lindsay, B. G., & Li, R. (2012). Local Modal Regression. *Journal of Nonparametric Statistics*, *24*(3), 647–663. https://doi.org/10.1080/10485252.2012.678848.

Zhu, Y., & Timmermann, A. (2020). *Can Two Forecasts Have the Same Conditional Expected Accuracy?*. https://doi.org/10.48550/arXiv.2006.03238.

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(10), 28–40 https://doi.org/10.59139/ws.2024.10.2
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 19.01.2024, poprawiony/revised: 10.09.2024, zaakceptowany/accepted: 20.09.2024

Measuring and assessing the health of refugees from Ukraine in Poland: quantitative-qualitative mixed-methods approach

Marek Cierpiał-Wolan,[a] Paloma Cuchí,[b] Federico Gerardo de Cosio,[c] Merkur Beqiri,[d] Rifat Hossain,[e] Dominik Rozkrut[f]

Abstract. Following the invasion of the Russian Federation on Ukraine on 24th February 2022, Poland has received an unprecedented influx of refugees. Statistics Poland and the World Health Organization (WHO) immediately recognized the urgent need for reliable data on the health needs of refugees and the barriers relating to their access to medical care, which would enable an effective response to the situation. The sudden surge in healthcare demand caused by the influx of those fleeing war required targeted interventions to ensure the well-being of both the refugees and the Polish population.

The aim of the article is to present the methodology and key findings of the "Health of Refugees from Ukraine in Poland" survey. This comprehensive study included questionnaires conducted among members of refugee households, supplemented by behavioral insights research based on in-depth interviews with refugees from Ukraine. Additionally, interviews were conducted with Polish and Ukrainian healthcare providers, authorities responsible for rescue operations and representatives of non-governmental organizations. The mixed-methods approach was applied, which encompasses both quantitative and qualitative research. Its flexibility and the use of data from administrative registers along with big data enabled real-time analyses and a more precise understanding of the refugees' health situation, with particular regard to the needs of women and older people.

The findings of the research highlight the importance of addressing language and financial barriers, issues related to the unfamiliarity with the health system, as well as service constraints (e.g. long waiting lists) and cultural barriers (e.g. stigma associated with mental health issues). The universal application of this approach has also been approved by the UN Statistical Commission, which recommended its further development and implementation in other countries facing a refugee crisis.

Keywords: refugees, refugee health, health needs, health access, healthcare, mixed methods, behavioral insights

JEL: C83, I14, I18

[a] University of Rzeszów, College of Social Science, Institute of Economics and Finance; Statistical Office in Rzeszów, Poland. ORCID: https://orcid.org/0000-0003-2672-3234. Corresponding author, e-mail: m.cierpial-wolan@stat.gov.pl.

[b] World Health Organization, Country Office Poland, Poland (formerly). ORCID: https://orcid.org/0000-0003-2516-0214. E-mail: cuchipal@outlook.com.

[c] World Health Organization, Country Office Poland, Poland (formerly). ORCID: https://orcid.org/0000-0003-4616-7212. E-mail: decosiog@icloud.com.

[d] World Health Organization, Country Office Poland, Poland (formerly). ORCID: https://orcid.org/0009-0005-7483-8126. E-mail: merkur@gmail.com.

[e] World Health Organization, Headquarters, Switzerland. ORCID: https://orcid.org/0000-0002-8542-1466. E-mail: hossainr@who.int.

[f] University of Szczecin, Faculty of Economics, Finance and Management, Institute of Economics and Finance; Statistics Poland, Poland. ORCID: https://orcid.org/0000-0002-0949-8605. E-mail: d.rozkrut@stat.gov.pl.

Pomiar i ocena stanu zdrowia uchodźców z Ukrainy w Polsce – podejście ilościowo-jakościowe oparte na metodach mieszanych

Streszczenie. Konsekwencją zbrojnej agresji Rosji na Ukrainę 24 lutego 2022 r. było przyjęcie przez Polskę bezprecedensowej liczby uchodźców. Główny Urząd Statystyczny i Światowa Organizacja Zdrowia (WHO) natychmiast dostrzegły konieczność szybkiego uzyskania wiarygodnych danych na temat potrzeb zdrowotnych uchodźców i barier związanych z dostępem do opieki medycznej, co umożliwiłoby skuteczne działanie w zaistniałej sytuacji. Nagły wzrost zapotrzebowania na usługi medyczne, spowodowany napływem osób uciekających przed wojną, wymagał bowiem ukierunkowanych interwencji, aby zapewnić dobrostan zarówno uchodźców, jak i ludności polskiej.

Celem artykułu jest przedstawienie metodologii i kluczowych wyników kompleksowego badania „Stan zdrowia uchodźców z Ukrainy w Polsce". Obejmowało ono badanie ankietowe przeprowadzone wśród członków gospodarstw domowych uchodźców, uzupełnione badaniem behawioralnym opartym na pogłębionych wywiadach z ukraińskimi uchodźcami. Ponadto przeprowadzono wywiady z polskimi i ukraińskimi przedstawicielami opieki zdrowotnej, organów odpowiedzialnych za działania ratunkowe i organizacji pozarządowych. Zastosowano metodę mieszaną uwzględniającą badania ilościowe i jakościowe. Jej elastyczność oraz wykorzystanie danych z rejestrów administracyjnych i big data umożliwiły przeprowadzenie analiz w czasie rzeczywistym i dokładniejsze zrozumienie potrzeb zdrowotnych uchodźców, ze szczególnym uwzględnieniem kobiet i osób starszych.

Badanie uwidoczniło potrzebę pokonywania barier językowych i finansowych, jak również wynikających z nieznajomości systemu opieki zdrowotnej i związanych z ograniczeniami w dostępie do usług medycznych (m.in. długim oczekiwaniem na wizytę lekarską) czy barier kulturowych (np. opór przed korzystaniem z usług psychiatrycznych z powodu stygmatyzacji chorych). Zastosowane podejście zostało zaaprobowane przez Komisję Statystyczną ONZ, która zaleciła jego dalszy rozwój i wykorzystywanie w innych krajach mierzących się z kryzysem uchodźczym.

Słowa kluczowe: uchodźcy, zdrowie uchodźców, potrzeby zdrowotne, dostęp do służby zdrowia, opieka zdrowotna, metody mieszane, badania behawioralne

1. Introduction

The literature review on refugee health often emphasizes that high-quality data are not always available, yet essential to guiding national and international health policies towards developing and implementing interventions required to address refugees' health needs effectively (Seagle et al., 2020). According to the World Health Organization (WHO), the main problems refugees encounter in a host country include not only health challenges, but also the fear of detention or deportation resulting from their precarious legal status, acts of discrimination, as well as social, cultural, linguistic, administrative and financial barriers (WHO, 2022c). The large influx of refugees poses a challenge to the host country with regards to their widely-understood integration into the society (Campomori et al., 2023); on the other hand, it presents an opportunity to demonstrate its commitment to humanitarian aid and

expand its cultural diversity through the successful integration of refugees. The host country's healthcare system, already strained by the aftermath of the COVID-19 pandemic, faced with a shortage of healthcare workforce and treatment-centric approaches (Campomori et al., 2023), becomes further challenged in terms of its ability to provide quality care to both the host population and the refugees (Seagle et al., 2020). Nevertheless, it also shows the country's capacity to adapt and provide quality care for all. One of the opportunities that refugees bring is the potential to contribute their skills and talents to the workforce and local communities. While there may be initial challenges related to job competition and reliance on public services (Fajth et al., 2019; Nie, 2015), refugees have the potential to become valuable contributors to the economy of the host country. Refugees also offer diverse perspectives and experiences that can enhance the social and cultural fabric of the host country.

Refugees may face unique health-related challenges connected with, for example, health disparities among their populations and barriers to reaching general healthcare or mental healthcare services for those dealing with the effects of forced displacement. The host country's healthcare system may struggle to meet the increased demand, resulting in limited resources and reduced access to care for all. This situation is likely to deepen the health inequalities and decrease the quality of care. However, it also presents an opportunity to improve the system's resilience and adaptability, potentially leading to long-term benefits for all.

In order for the complex process of refugee integration to be successful, particularly in terms of the transition from reception to self-sufficiency, a coordinated response from all the stakeholders, including state entities (such as health, social, labor and legal sectors) and non-state actors, like non-governmental organizations (NGOs) and the refugees themselves, is required (Campomori et al., 2023). This multifaceted collaboration, when supported by reliable data, can facilitate a smoother and more successful transition for those seeking refuge, leading to positive outcomes for both refugees and the host community. The experience of Poland in generating this information, as detailed in this study, offers valuable insights that can strengthen integration efforts made in other countries.

Through the careful examination of the healthcare experiences of refugees from Ukraine in Poland, this study contributes to the existing body of knowledge on refugee health and provides valuable insights for policymakers and practitioners working to improve healthcare access and utilization among refugee populations (Seagle et al., 2020).

Since the start of the full-scale war following the Russian Federation's invasion in February 2022, the influx of refugees into Poland has been unprecedented in the recent times and has reached the Second World War level of population movement. The number of people crossing the border between Ukraine and Poland (to stay or move on to other countries) has reached over 14.5 million, representing about 27.5%

of the total estimated population of Poland. As of August 2023, around 12 million individuals crossed the Polish-Ukrainian border in both directions. Among those who stayed in Poland, over 1.6 million refugees have been registered in Poland and received a PESEL identification number (Personal Identification Number in the Common Electronic System of Population Register; Dane.gov.pl, 2023; Sas, 2023); this number is equivalent to approximately 4.4% of Poland's population (37.6 million), suggesting that Poland has successfully integrated these refugees into its socio-economic environment within a short period of time. Each person with a PESEL number is entitled to public healthcare, parental benefits, family care and social assistance (Sas, 2023). This great influx of individuals to Poland has placed considerable demands on the healthcare system, which had already been experiencing the strain caused by the COVID-19 pandemic and systemic issues. Even with Poland's dedicated inclusive healthcare service delivery to refugees, the challenge is to plan a strategic response to meet an uncertain level of need for health services without health status and needs representative data, while ensuring that the host population's access to quality healthcare remains unhindered (United Nations Statistical Commission [UNSC], n.d.).

In response, Statistics Poland, the WHO Country Office in Poland, the Behavioural and Cultural Insights Unit at the WHO Regional Office for Europe and the Health and Migration Programme of the WHO Headquarters conducted a study called "Health of Refugees from Ukraine in Poland". It is based on quantitative data collected from refugee households[1] and research on qualitative behavioral insights. This mixed-methods approach was used to gain insights into refugees' health status, their perceived health needs and the barriers they faced when accessing healthcare services (Spiegel, 2022; UNSC, n.d.; WHO & Statistics Poland, 2023). The study also explored the healthcare providers' perspectives on the services offered to refugees from Ukraine.

The comprehensiveness of data produced by this mixed-methods approach provided depth and breadth of information on the healthcare needs of the refugees. This approach serves as an evidence-informed model for rigorous data collection that provides both statistical evidence from a representative sample and a deeper understanding of the perceived factors that facilitate or hinder access to healthcare among the refugee population.

This study provided an important opportunity to establish a strengthened, dynamic, collaborative partnership between Statistics Poland, the Polish Ministry of Health, WHO and other pivotal stakeholders, including donor and multilateral agencies. This collective effort bolstered coordination between institutions that collect data and those which use them to enable the execution of robust, effective and coherent strategies.

[1] In the study, the term "household" refers to a group of individuals who meet the following criteria: they traveled together from Ukraine, left Ukraine due to armed actions by the Russian Federation, and plan to live together during their stay outside of Ukraine.

2. Methods and data sources

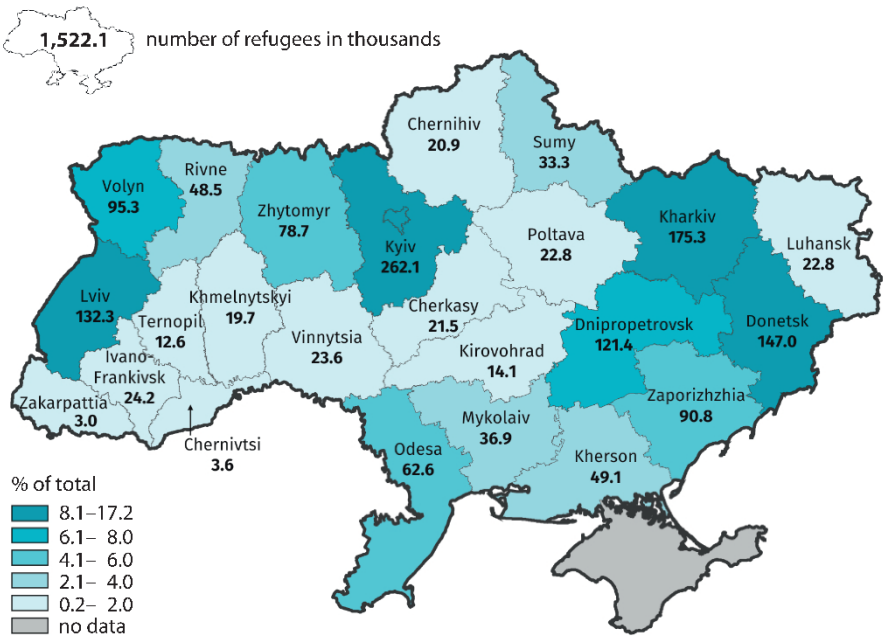
This article describes the application of the mixed-methods approach, which integrates both qualitative and quantitative analyses to ensure a comprehensive understanding of the research topic. Mixed methods, which combine the numerical precision of quantitative data with in-depth insights obtained from qualitative research, allow capturing a more holistic view of complex issues (Creswell & Plano Clark, 2018). The application of the mixed-methods approach resulted in a collection of statistical data from a representative sample, offering broad generalizability and enabling a nuanced understanding of the contexts and barriers related to access to healthcare among refugees. By leveraging the strengths of both methodologies, the study not only delivers a more complete picture of the health-related needs but also facilitates more precise and actionable conclusions and recommendations.

The quantitative analyses involved a two-stage stratified random sampling method, with the sampling conducted at the voivodship level. In the first stage, locations were selected based on simple random sampling; in the second stage, people living in these locations were chosen for the research using systematic sampling. The survey questionnaire was crafted in Ukrainian, English, and Polish. Statistical interviewers carried out questionnaires in 94 locations in Podkarpackie and Lubelskie Voivodships (administrative division units in Poland corresponding to NUTS 2 regions in the European Union), obtaining 1,800 completed questionnaires. The collected data focused on demographic characteristics, health status and needs, disease prevention and vaccination, mental health, health costs and access to healthcare within the surveyed population (WHO & Statistics Poland, 2023). The survey results were estimated using sample weights calculated on the basis of the number of registered refugees from Ukraine.

The importance of big data as a source of information is growing along with its constantly increasing level of accessibility. Big data sources were used to calibrate the weights to reflect the true pattern of refugee movements. The further development of the "Health of Refugees from Ukraine in Poland" survey is planned using data from big data sources, i.e. mobile phone and payment card operators. This method is likely to be useful in obtaining information on the movement of refugees and to monitor expenditures related to healthcare.

It is important to highlight that the survey included individuals from all regions of Ukraine, which was crucial for ensuring the credibility of the study. Figure 1 shows the number of refugees from each Ukrainian region who found temporary shelter in Poland between 24th February and 31st August 2022 and who were residing in Poland at the time of the survey.

Figure 1. Ukrainian refugees by place of origin



Source: WHO and Statistics Poland (2023).

The "Health of Refugees from Ukraine in Poland" survey was preceded by a pilot study[2] conducted at the reception points on the Polish-Ukrainian border in Podkarpackie Voivodship. At the end of May 2022, the activities of most reception points at the border were suspended as the number of people crossing the border decreased. As many refugees continued to move through Poland, it became necessary to identify new refugee locations for the survey.

In tandem, a behavioral insights study, supported by the WHO Regional Office for Europe, delved into the health service needs and access for Ukrainian refugees in Poland. This study employed a modified capability, opportunity, motivation and behaviors (COM-B) framework to explore the barriers and enablers of behaviors. The research questions (WHO & Statistics Poland, 2023) aimed to identify:

- health-related service needs and expectations of refugees from Ukraine in Poland in the area of prevention, treatment, care and previous health-seeking behaviors;
- barriers and facilitators in accessing and utilizing healthcare services, focusing on: the awareness and knowledge of the available services and treatment needs, the motivational factors and barriers to seeking care, the perceived opportunities and experiences with health services, including any stigma or discrimination and the level of the support received from family, community and health services;
- behavioral and cultural factors influencing health service uptake.

[2] The precision for the study was 98% and 97% for the pilot study.

After the ethical approvals were secured, participants aged 18 and above, who had left Ukraine due to the war and resided in Poland for at least two weeks, were purposively selected through maximum variation sampling. In-depth online interviews, conducted by the Ukrainian research agency Sociologist, involved a total of 35 participants drawn from the 1,800 households surveyed between 25th August and 7th September 2022. These studies yielded valuable insights contributing to the formulation of programs and services aimed at assisting refugees from Ukraine in Poland in addressing their health-related concerns and issues.

Additionally, 25 interviews were conducted with Polish and Ukrainian healthcare providers, response authorities and staff from NGOs. The objective was to glean insights into their perspectives on the challenges associated with providing healthcare to refugees. Participants were selected through purposive sampling to ensure a diverse range of expertise and experiences relevant to the study's focus.

3. Results

The findings from the relevant components of the study (WHO & Statistics Poland, 2023) provided valuable insights that were used to actively shape the programs and services designed to assist refugees from Ukraine in Poland. It was anticipated that most refugees would be women and children, given that the Ukrainian government had restricted the departure of men under the age of 60, excluding cases of exceptional family circumstances. The results of the quantitative survey of 1,800 households were that 70% (1,260) of refugees were females, 59% (1,062) of refugees were over 18 years of age, and 41% (738) were children under the age of 18. With regards to the level of education, 53% had higher post-primary education, with 47% holding university degrees. The top five health problems reported were acute illness (43%), chronic disease (39%), oral health conditions (18%), mental health conditions (6%) and COVID-19-related illnesses (3%). The main identified healthcare access challenges were associated with language barriers and limited information about the functioning of the health system (50%), which also hindered refugees' ability to explain their symptoms to healthcare providers, and the health cost (33%), with one in four respondents spending over 25% of their income on healthcare. Other challenges related to difficulties in reaching medical facilities (14%), long waiting time (7%) and the unavailability of medical care or treatment (7%).[3] Out of the total of refugees who arrived in Poland, 35% were vaccinated against COVID-19. This is similar to the proportion of those vaccinated in Ukraine; according to WHO data, those fully vaccinated accounted for 37% of Ukraine's total population as of 27th February 2022.

[3] The percentages add up to over 100% as respondents were able to give more than one answer.

These figures are lower than in Poland, where the percentage of people who were fully vaccinated reached 60% of the total population in the same or similar period.

Thanks to the integration of data from both quantitative and qualitative surveys, which includes big data as well as administrative records, significant differences in certain characteristics compared to administrative records have been identified. Consequently, it was found that not all refugees registered to obtain the PESEL number. This conclusion was confirmed by the "Health of Refugees from Ukraine in Poland" survey, which revealed that across every age group and for both men and women, the number of refugees was higher than the number recorded in the PESEL register (see Figure 2). The largest discrepancy was observed among the population over the age of 54.

Figure 2. Number of Ukrainian refugees in Poland by gender in each age group: household survey and the PESEL register

Source: WHO and Statistics Poland (2023).

The qualitative study (WHO & Statistics Poland, 2023) provided detailed insights that complemented the statistical data and offered a deeper understanding of the healthcare needs of the refugees as well as the barriers to and drivers of accessing care. For example, the interviews helped uncover gender-specific dimensions of the barriers to seeking and receiving timely care, such as women encountering challenges related to childcare and caregiving to older adults without the usual social support mechanisms. The qualitative research also allowed a more nuanced understanding of

the sensitivity related to mental health. Quantitative survey interviewers struggled to elicit responses from people on this topic, and the qualitative study showed that mental health issues tend to be stigmatized, which, as some respondents suggested, may be related to the history of mental healthcare in former Soviet states (Ociepa-Kicińska & Gorzałczyńska-Koczkodaj, 2022). In addition, practical issues were found to impact uptake of mental health services, such as the type of services offered (group versus individual), location and language. Moreover, during the in-depth interviews, dental care was mentioned as a priority need and this was confirmed as a high-priority issue in the representative sample of refugees in the quantitative survey. Despite facing some barriers to accessing health services, refugees in Poland had an overall positive perception of the healthcare system.

In-depth interviews with both Polish and Ukrainian healthcare providers working in Poland, the response authorities and representatives of NGOs offered additional and important perspectives. The results of these interviews show that healthcare workers and NGO staff are motivated by compassion for their neighbors and that NGOs play an important role in increasing access to health services. Polish healthcare providers confirmed that language is a barrier when interacting with refugees, as are the other challenges that the users of the Polish healthcare system encounter, such as long waiting time for appointments.

The results of the household survey and the behavioral insights research helped identify these problems and elicited refugees' expectations about the healthcare system. Based on the findings of these surveys, there is a need for information targeting older people, those with disabilities, those suffering from chronic diseases or showing low health literacy. Moreover, targeted information should be disseminated among refugees on prevention services, specialized care and vaccinations. The findings from both surveys also indicate a need for mental health services that factor in the stigma that may prevent people from Ukraine from seeking help.

The described data collection strategy combined quantitative and qualitative data to provide evidence for a nuanced understanding of the healthcare situation of the refugees and to allow for informed decision-making by national authorities. This holistic perspective enabled stakeholders responsible for refugee health in Poland to design interventions that effectively addressed refugees' immediate healthcare needs, while also allowing more resilient and inclusive healthcare responses to meet the broader needs of refugees and the host community.

On the basis of the results of the fruitful collaboration between Statistics Poland, national health sector agencies and entities from the three levels of the WHO, the following considerations are noteworthy:

- high-quality and timely data are needed. Data on the healthcare needs facing refugees and migrants and the barriers to accessing healthcare services are often scarce and of poor quality. Reliable, robust and timely data are essential for a better understanding of population movements and necessary to formulate appropriate healthcare-related responses;

- in order to improve relevant health-related policies, standardized tools are needed, i.e. systematically collected and disaggregated by migratory status data, representative of both refugee and migrant populations, comparable across countries and over time. Disaggregated data, including those on refugees and migrants, allow the healthcare stakeholders to understand and address the health needs of refugees, develop inclusive public health approaches, track the progress towards the national and global health goals, and enable decision-makers to understand and respond to public health challenges at a national level.

Achieving many of the health and health-related Sustainable Development Goals and targets will require fit-for-purpose data on the health of refugees and migrants. Data integrated from various sources create a more comprehensive picture of the examined situation. In addition to the quantitative and qualitative methods used, exploring data integration methods using data from the health, social, economic and other sectors are likely to prove useful. Big data and other innovative and emerging sources of data can also contribute to the broader understanding of the refugee health-related issues. The collaboration of various entities and the coordination of their activities are key to making progress and achieving success in the area of health and data, as demonstrated by the cooperation between Statistics Poland and WHO during the course of preparing the "Health of Refugees from Ukraine in Poland" study. Extending this type of collaboration across the health sector and national statistical systems ensures a timely and effective collection and availability of data on the health of refugees and migrants around the world, benefiting the health of more than a billion people worldwide.

4. Conclusions

The rapid and vast influx of refugees from Ukraine into Poland and other neighboring countries has placed a substantial burden on their healthcare systems; it has also prompted these systems to adapt and innovate, showcasing their resilience in the face of an unprecedented challenge. While the surge in demand for healthcare has undoubtedly placed additional pressure on the existing resources, it has also highlighted the importance of robust health information systems and data-driven decision-making. This influx has prompted a critical examination of the healthcare capacity and a renewed focus on the development of innovative solutions to meet the diverse needs of both the host population and the refugees. By leveraging timely and accurate data, healthcare systems can proactively address challenges, optimize resource allocation and ensure access to quality care for all. This proactive approach, informed by comprehensive data and collaborative efforts, can mitigate the potential strain and foster a more efficient and effective healthcare response.

The surveys conducted by WHO and Statistics Poland (2023), based on a mixed-methods approach that included both quantitative and qualitative research, revealed new social phenomena and processes. Common health issues included acute and

chronic illnesses, oral health problems, mental health concerns and COVID-19. Key difficulties in healthcare access were caused by language barriers and challenges in explaining symptoms, the limited understanding of the functioning of the health system, health costs, long waiting time as well as the benefits available for the refugees. As a result, not all refugees are able to take full advantage of the existing healthcare resources and services. The qualitative research highlighted additional gender-specific obstacles, mental health stigma and the critical role of NGOs. Overall, refugees had a positive perception of the Polish health system despite these barriers.

Statistical offices and national health sector agencies complement each other as producers and users of data. Their collaboration proves essential when a timely collection, analysis and use of data for effective decision-making are needed. Joint planning, data sharing and capitalizing on each other's comparative advantages are key to support evidence-based decision-making and policy formulation in the health sector, ultimately leading to better health-related outcomes and the well-being of the population.

A distinct synergy effect emerges from the collaboration between statistical offices and national health sector agencies, benefitting both sides. Firstly, for WHO reports, statistical offices can provide valuable expertise in data collection, analysis and quality assurance that can provide the health sector with access to reliable and comparable data for health indicators. On the other hand, the health sector data from administrative records, surveys, registries and vital statistics can enrich the statistical databases and, most importantly, enhance the understanding of the health context and the interpretation of data. The collaboration between these institutions helps ensure that the health-related data are accurate, complete, consistent and timely.

Secondly, the cooperation between statistical offices and the health sector enhances the coordination aimed at strengthening health information systems and makes them more fit-for-purpose. By collaborating with statistical offices, the health sector can adopt common frameworks and tools that facilitate the interoperability and comparability of health data across different sources and at various levels. This is paramount to consolidating the efforts for cost-effective and robust evidence-based actions.

Finally, the collaboration between statistical offices and the health sector supports evidence-based decision-making and policy formulation in the health sector. A practical example of this is the study presented herein, which brings together statistical offices, health ministries, development partners, civil society organizations and academic institutions to improve the availability and use of health data for better health outcomes.

It should be noted that big data, which is a vast and real-time source of information, provides an opportunity to implement actions that align with the evolving situations and optimize the coordination of the available resources. The collaboration between health and statistics in the form of the mixed-methods approach integrated with administrative and big data to cover the information gap in Poland has paved the way for the exploration of big data strategies to gather and generate information in near real-time.

The results of this survey and the use of the mixed-methods approach were presented at a side event during the 54th session of the UNSC, "Towards a global measurement framework of health of refugees and migrants: lessons learned from a refugee health survey". The chief statisticians of the world noted "the initial work undertaken on measuring health of refugees" and called on WHO and partners to further develop this work and collect transformative data, ensuring alignment with existing definitions and statistical frameworks on refugees and migrants, and coordination with Expert Group on Migration Statistics (UNSC, n.d.). These efforts, jointly with the extension of WHO's global action plan to 2030 and the emphasis on behavioral science and cultural insights by WHO's Executive Board and Regional Committee for Europe, have the potential to significantly improve the delivery of health services to refugees (WHO, 2022a, 2022b, 2023).

The study described in this article represents a significant advancement in developing innovative approaches to collecting and analyzing data on refugee health. The mixed-methods approach as well as big data and administrative registers used to calibrate the weights, underscore the need for the continuous adaptation of methodological frameworks in response to the evolving challenges posed by the ongoing war in Ukraine. The flexibility and responsiveness of this methodology allow for real-time updates and offer a more precise understanding of refugees' health needs, thereby facilitating timely and effective policy responses. Furthermore, the universal applicability of this approach allows its implementation in other countries facing refugee crises worldwide. This methodology can be adapted to various contexts, providing a valuable tool for improving health outcomes for displaced populations worldwide. These innovations highlight the importance of dynamic, evidence-based solutions in the face of complex humanitarian crises.

Acknowledgements

We extend our gratitude to Martha Scherzer for her expert contribution on the behavioral insights component of this project, which significantly enhanced our work.

References

Campomori, F., Casula, M., & Kazepov, Y. (2023). Understanding social innovation in refugee integration: actors, practices, politics in Europe. *Innovation: The European Journal of Social Science Research*, *36*(2), 158–170. https://doi.org/10.1080/13511610.2023.2211893.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3rd ed.). SAGE Publications.

Dane.gov.pl. (2023, August 8). *Registered Applications for Granting UKR Status in Connection with the Conflict in Ukraine.*

Fajth, V., Bilgili, Ö., Loschmann, C., & Siegel, M. (2019). How do refugees affect social life in host communities? The case of Congolese refugees in Rwanda. *Comparative Migration Studies*, *7*(1), 1–21. https://doi.org/10.1186/s40878-019-0139-1.

Guziak, M., & Bastrzyk, Z. (2023). Sektor ochrony zdrowia w obliczu konfliktu zbrojnego. *Rocznik Bezpieczeństwa Międzynarodowego*, *17*(1). https://doi.org/10.34862/rbm.2023.1.3.

Lewtak, K., Kanecki, K., Tyszko, P., Goryński, P., Kosińska, I., Poznańska, A., Rząd, M., & Nitsch-Osuch, A. (2022). Hospitalizations of Ukrainian Migrants and Refugees in Poland in the Time of the Russia-Ukraine Conflict. *International Journal of Environmental Research and Public Health*, *19*(20), 1–10. https://doi.org/10.3390/ijerph192013350.

Nie, Z. (2015, September 16). *The Effects of Refugees on Host Countries*. https://globaledge.msu.edu/blog/post/30996/the-effects-of-refugees-on-host-.

Ociepa-Kicińska, E., & Gorzałczyńska-Koczkodaj, M. (2022). Forms of Aid Provided to Refugees of the 2022 Russia–Ukraine War: The Case of Poland. *International Journal of Environmental Research and Public Health*, *19*(12), 1–17. https://doi.org/10.3390/ijerph19127085.

Sas, A. (2023, January). *Refugees from Ukraine with an assigned PESEL number in Poland 2022–2023.* https://www.statista.com/statistics/1301649/poland-refugees-from-ukraine-with-an-assigned-pesel-by-region/.

Seagle, E. E., Dam, A. J., Shah, P. P., Webster, J. L., Barrett, D. H., Ortmann, L. W., Cohen, N. J., & Marano, N. N. (2020). Research ethics and refugee health: a review of reported considerations and applications in published refugee health literature, 2015-2018. *Conflict and Health*, *14*, 1–15. https://doi.org/10.1186/s13031-020-00283-z.

Spiegel, P. B. (2022). Responding to the Ukraine refugee health crisis in the EU. *The Lancet*, *399*(10341), 2084–2086. https://doi.org/10.1016/S0140-6736(22)00841-8.

United Nations Statistical Commission. (n.d.). *Expert Group on Refugee, Internally Displaced Persons and Statelessness Statistics*. Retrieved September 18, 2023, https://unstats.un.org/UNSDWebsite/statcom/session_54/documents/BG-4e-EGRISS-202022Meeting-E.pdf.

Wojtkowiak, M. (2022). *General characteristics of possible consequences of trauma and forms of aid granted to Ukrainian children refugees*, *10*, 151–167. Labor et Educatio. https://doi.org/10.4467/25439561LE.22.012.17538.

World Health Organization. (2022a). *EB152/36 Global strategies and plans of action that are scheduled to expire within one year. WHO global action plan on promoting the health of refugees and migrants, 2019–2023*. https://apps.who.int/gb/ebwha/pdf_files/EB152/B152_36-en.pdf.

World Health Organization. (2022b). *European regional action framework for behavioural and cultural insights for health, 2022–2027*. https://iris.who.int/bitstream/handle/10665/360898/72wd06e-rev1-RegActionFramework- BCI-220516?sequence=1&isAllowed=y.

World Health Organization. (2022c). *World report on the health of refugees and migrants*. World Health Organization. https://iris.who.int/handle/10665/360404.

World Health Organization. (2023). *Behavioural sciences for better health. Draft decision proposed by Bangladesh, Brunei Darussalam, Jamaica, Japan, Malaysia, Philippines, Qatar, Singapore, Slovakia, South Africa, Thailand and United States of America*. https://apps.who.int/gb/ebwha/pdf_files/EB152/B152_CONF6-en.pdf.

World Health Organization, & Statistics Poland. (2023). *Health of refugees from Ukraine in Poland 2022. Household survey and behavioural insights research.* https://stat.gov.pl/files/gfx/portalinformacyjny/pl/defaultaktualnosci/6377/7/1/1/raport__who_21.02.pdf.

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(10), 41–58 https://doi.org/10.59139/ws.2024.10.3
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 12.07.2023, poprawiony/revised: 15.02.2024, zaakceptowany/accepted: 30.09.2024

Czy warto budować modele prognozowania upadłości przedsiębiorstw uwzględniające kryterium przynależności sektorowej?[1]

Sergiusz Herman[a], Bartłomiej Lach[b]

Streszczenie. Prognozowanie upadłości przedsiębiorstw jest bardzo ważnym zagadnieniem, stanowiącym przedmiot zainteresowania szerokiej grupy interesariuszy. Z tego powodu od początku XX w. nieprzerwanie podejmowane są prace nad rozwojem skutecznych narzędzi wykorzystywanych w tym kierunku. Głównym celem badania omawianego w artykule jest zweryfikowanie zasadności konstruowania sektorowych modeli prognozowania upadłości przedsiębiorstw. Badanie zostało przeprowadzone na podstawie danych finansowych 800 polskich przedsiębiorstw za lata 2015–2019, pobranych z bazy EMIS Professional. W analizie uwzględniono sześć metod doboru zmiennych, sześć metod uczenia maszynowego oraz 500 różnych losowych prób uczących i testowych.

Uzyskane wyniki prowadzą do wniosku, że determinanty upadłości w poszczególnych sektorach gospodarki są odmienne. Modele opierające się na grupie przedsiębiorstw prowadzących jednorodną działalność gospodarczą pozwalają uzyskać przeciętnie wyższe wskaźniki trafności klasyfikacji. Na tej podstawie można stwierdzić, że konstruowanie takich modeli jest zasadne.

Słowa kluczowe: przedsiębiorstwo, upadłość, przynależność sektorowa, uczenie maszynowe
JEL: G33, C38, C53

Is it worth building models for forecasting bankruptcy of enterprises taking into account the criterion of sectoral affiliation?

Abstract. Forecasting bankruptcy of enterprises is a very important area, and it is of interest to a wide group of stakeholders. For this reason, since the beginning of the last century, continuous work aimed at developing effective tools used to this end has been undertaken. The main goal

[1] Artykuł został opracowany na podstawie referatu wygłoszonego na XVI Międzynarodowej Konferencji Naukowej im. Profesora Aleksandra Zeliasia „Modelowanie i prognozowanie zjawisk społeczno-gospodarczych", która odbyła się w dniach 8–11 maja 2023 r. w Zakopanem. / The article is based on a paper delivered at the 16th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena, held on 8–11 May 2023 in Zakopane, Poland.

[a] Uniwersytet Ekonomiczny w Poznaniu, Instytut Informatyki i Ekonomii Ilościowej, Polska / Poznań University of Economics and Business, Institute of Informatics and Quantitative Economics, Poland.
ORCID: https://orcid.org/0000-0002-2753-1982. Autor korespondencyjny / Corresponding author,
e-mail: sergiusz.herman@ue.poznan.pl.

[b] Analyx sp. z o.o. sp. k., Polska / Poland. ORCID: https://orcid.org/0000-0002-2023-0378.
E-mail: lach.bartlomiej@gmail.com.

of the study presented in this article is to verify the legitimacy of constructing industry models for forecasting the bankruptcy of enterprises. The research was conducted on the basis of financial data of 800 Polish companies for the years 2015–2019, drawn from the EMIS Professional database. The analysis used six methods of variable selection, six machine learning methods and 500 distinct random training and testing samples.

The obtained results lead to the conclusion that there are different determinants of bankruptcy in different sectors of the economy. Models constructed on the basis of a group of enterprises conducting homogeneous economic activity make it possible to achieve, on average, higher classification accuracy rates, which demonstrates that the construction of such models is justified.

Keywords: enterprise, bankruptcy, sectoral affiliation, machine learning

1. Wprowadzenie

Upadłość przedsiębiorstw jest immanentną cechą każdej gospodarki rynkowej. W ten sposób rynek eliminuje przedsiębiorstwa niewypłacalne, niedziałające efektywnie. Zjawisko to, ze względu na swą powszechność, stanowi przedmiot licznych badań naukowych. Ich zasadniczym celem jest określenie poziomu ryzyka, że przedsiębiorstwo zostanie dotknięte problemami finansowymi, które nie pozwolą mu dalej funkcjonować. Pierwsze wielowymiarowe narzędzia do prognozowania upadłości przedsiębiorstw zostały zaproponowane w latach 60. XX w. przez Altmana (1983). Szeroką gamę badań można podzielić zgodnie z podstawowymi kierunkami ich prowadzenia. Przeważająca większość dotyczy poszukiwania coraz bardziej sprawnych, zaawansowanych matematycznie metod klasyfikacyjnych. Zakres stosowanych metod jest bardzo szeroki – od statystycznych metod klasyfikacyjnych przez zaawansowane metody uczenia maszynowego po bardzo popularne w ostatnich latach metody hybrydowe (Jabeur i in., 2022; du Jardin, 2018). Początkowo modele upadłości były konstruowane przede wszystkim z wykorzystaniem odmiennych wskaźników finansowych (Mroczkowska i Rogowski, 2010). Z czasem ich lista została rozszerzona o wskaźniki rynkowe (Campbell i in., 2008), wskaźniki dotyczące ładu korporacyjnego (Chan i in., 2016; Liang i in., 2016), wskaźniki dotyczące innowacyjności firm (Bai i Tian, 2020) oraz zmienne odpowiedzialne za położenie geograficzne (Gabbianelli, 2018; Ptak-Chmielewska, 2018). Przedmiotem analiz są też często wpływ zbilansowania próby badawczej na uzyskiwane wyniki (Veganzones i Séverin, 2018; Zhang i in., 2022; Zhou, 2013), obecność wartości odstających (Nyitrai i Virág, 2019) czy metoda doboru zmiennych prognostycznych (Liang i in., 2015; Zoričák i in., 2020).

Głównym celem badania omawianego w artykule jest zweryfikowanie zasadności konstruowania sektorowych modeli prognozowania upadłości przedsiębiorstw, a celami pobocznymi są zbadanie zależności pomiędzy wielkością próby badawczej a trafnością klasyfikacji wykorzystanych metod oraz porównanie sposobu działania

klasyfikatorów zbudowanych z wykorzystaniem różnych metod. Na zasadność podjęcia takiego badania może wskazywać to, że przedsiębiorstwa z różnych sektorów gospodarki muszą się mierzyć z odmiennym poziomem konkurencji (Chava i Jarrow, 2004). Diametralnie odmienny charakter prowadzonej działalności wpływa także na strukturę aktywów i pasywów firm, a to z kolei ma odzwierciedlenie w sektorowym zróżnicowaniu poziomu poszczególnych wskaźników finansowych. Świadczą o tym badania poświęcone kondycji finansowej przedsiębiorstw z różnych sektorów gospodarki (Dudycz i Skoczylas, 2023).

Realizacja celu badania omawianego w artykule wymagała porównania determinant upadłości przedsiębiorstw w różnych sektorach gospodarki oraz wskaźników trafności klasyfikacji uzyskanych dla modeli sektorowych i modelu ogólnego. Na podstawie przeglądu literatury można stwierdzić, że niewiele opracowań porusza tę problematykę.

2. Przegląd literatury

Konstrukcja modelu prognozowania upadłości przedsiębiorstw wymaga zgromadzenia danych finansowych dla określonej próby przedsiębiorstw. Już Altman (1983) zwrócił uwagę na to, że podmioty objęte badaniem powinny prowadzić możliwie najbardziej jednorodną działalność gospodarczą, co podkreśla wagę specyfiki sektorowej analizowanych podmiotów. Dzięki dostępowi do obszernych baz danych w literaturze światowej często podejmowane są próby konstrukcji modeli prognozowania upadłości przedsiębiorstw dla konkretnych sektorów gospodarki. Można tutaj wymienić np. badania związane z przewidywaniem upadłości w sektorze bankowym (Ravi Kumar i Ravi, 2007), budownictwa (Heo i Yang, 2014), przemysłowym (Chen i in., 2013), restauracji (Becerra-Vicario i in., 2020) czy w rolnictwie (Rajin i in., 2016). Bardzo niewielu autorów podejmuje się zweryfikowania, czy konstruowanie modeli sektorowych jest uzasadnione (Bellovary i in., 2007; Gill de Albornoz i Giner, 2013; Laguillo i in., 2019).

Pierwsze modele prognozowania upadłości przedsiębiorstw w Polsce nie uwzględniały rodzaju prowadzonej działalności (Gajdka i Stos, 1996; Hadasik, 1998; Hołda, 2001; Pogodzińska i Sojak, 1995) lub były konstruowane dla najliczniej reprezentowanego sektora firm produkcyjnych (Korol, 2005; Korol i Prusak, 2005; Pisula, 2017; Pociecha, 2014; Prusak, 2005). Wraz ze wzrostem dostępności danych finansowych zaczęto konstruować modele dla innych sektorów gospodarki. Można tu wspomnieć chociażby o modelach skonstruowanych dla sektorów rolnictwa (Boratyńska i Grzegorzewska, 2018; Grzegorzewska, 2011; Grzegorzewska i Runowski, 2008), budowlanego (Hołda 2009; Król i Stefański, 2014; Pisula, 2010a; Rusiecki i Białek-Jaworska, 2015), sektora bankowego (Appenzeller i Nowara, 2003),

deweloperów (Pisula, 2010b), leśno-drzewnego (Noga i in., 2014), usług spedycyjnych, transportowych i logistycznych (Brożyna i in., 2016; Juszczyk, 2010; Karbownik, 2013), spożywczego (Zielińska-Sitkiewicz, 2016) czy przemysłu mięsnego (Wysocki i Kozera, 2012).

W polskiej literaturze można spotkać próby konstruowania modeli dla kilku sektorów gospodarki (Herman, 2017; Hołda, 2006; Iwanowicz, 2018; Jagiełło, 2013; Juszczyk i Balina, 2014; Potoczna i Wiśniewska, 2013). Nieliczni badacze zdecydowali się zweryfikować, czy konstruowanie takich modeli prognozowania upadłości przedsiębiorstw jest uzasadnione. Hołda (2006) w swoim badaniu dysponował próbą 374 przedsiębiorstw, reprezentujących trzy sektory gospodarki. Zastosował cztery metody klasyfikacyjne i trzy metody doboru zmiennych. Zasadność konstrukcji modeli sektorowych zweryfikował poprzez porównanie ich zdolności prognostycznych dla dwóch prób testowych – próby spółek z tego samego sektora, dla której był szacowany model, oraz dla próby przedsiębiorstw prowadzących odmienną działalność. Autor zauważył, że łączenie podmiotów z różnych sektorów prowadzi do pogorszenia trafności uzyskanych prognoz.

Analogiczne badanie przeprowadzili Juszczyk i Balina (2014). Ich analiza, która również dotyczyła trzech sektorów gospodarki (reprezentowanych przez 180 spółek), składała się z dwóch części. Pierwsza polegała na wykorzystaniu znanych w literaturze polskiej i zagranicznej modeli prognozowania upadłości do zbadania trafności klasyfikacji dla poszczególnych sektorów gospodarki i dla całej dostępnej próby badawczej, druga – na skonstruowaniu własnych modeli prognostycznych z wykorzystaniem analizy dyskryminacyjnej: trzech sektorowych i jednego ogólnego (nieuwzględniającego specyfiki sektorowej firm). Porównując uzyskane wyniki dotyczące trafności klasyfikacji modeli, autorzy uznali, że konstruowanie modeli sektorowych może przyczynić się do poprawy jakości stawianych prognoz.

Podobnie Herman (2017) skonstruował modele prognozowania upadłości dla trzech sektorów gospodarki, tj. budownictwa, przetwórstwa przemysłowego i handlu. Próba badawcza, na której się opierał, obejmowała 180 spółek akcyjnych. Autor wykorzystał, jak wspomniani Juszczyk i Balina (2014), analizę dyskryminacyjną, a błędy predykcji oszacował z wykorzystaniem estymatora wielokrotnego repróbkowania. Uzyskane wyniki pozwoliły na wyciągnięcie wniosku, że modele ukierunkowane sektorowo nie charakteryzują się niższym błędem predykcji, a tym samym większą zdolnością predykcyjną w stosunku do wyników uzyskanych dla modelu ogólnego.

We wszystkich opisanych powyżej badaniach modele konstruowane i oceniane były tylko raz – na podstawie jednokrotnie wylosowanych prób uczącej i testowej. Ponadto nie zwrócono uwagi na fakt, że poszczególne modele były szacowane i testo-

wane na podstawie prób badawczych różnej wielkości. W literaturze często podkreśla się, jak istotna z punktu widzenia uzyskiwanych wyników klasyfikacji jest wielkość próby: wszystkie porównywane modele sektorowe powinny być budowane i testowane na podstawie tak samo dużej próby (Beleites i in., 2013; Vabalas i in., 2019). W przedstawionych analizach nie zweryfikowano także odporności uzyskanych wyników na zastosowane metody badawcze.

3. Metoda badania

Punktem wyjścia do przeprowadzenia badania omawianego w niniejszym artykule było zgromadzenie odpowiedniej próby badawczej obejmującej przedsiębiorstwa, wobec których sąd wydał postanowienie o ogłoszeniu upadłości (dalej określane jako *bankruci*), oraz przedsiębiorstwa funkcjonujące na rynku i będące w dobrej kondycji finansowej (dalej określane jako *podmioty zdrowe*). W tym celu wykorzystano bazę EMIS Professional[2]. Biorąc pod uwagę dostępność danych finansowych, wylosowano 400 przedsiębiorstw, wobec których wydano postanowienie o ogłoszeniu upadłości, a postępowanie upadłościowe wszczęto w latach 2016–2020. W analizie pominięto lata 2021–2022 ze względu na drastyczne zmiany warunków gospodarczych funkcjonowania firm oraz prawa upadłościowego i restrukturyzacyjnego w Polsce spowodowane pandemią COVID-19.

Wylosowane podmioty reprezentowały trzy sektory gospodarki: budownictwo (PKD 41.10–43.99Z), przetwórstwo przemysłowe (PKD 10.11–33.20Z) oraz handel hurtowy i detaliczny (PKD 46.11–47.99Z). Do tej próby dolosowano 400 podmiotów zdrowych. Wykorzystano próbę zbilansowaną, jak w zdecydowanej większości badań poświęconych prognozowaniu upadłości przedsiębiorstw. W przypadku próby niezbilansowanej uzyskana trafność klasyfikacji, szczególnie dla próby mniej licznej, często jest bowiem znacznie gorsza. Rozwiązanie tego problemu wymagałoby np. zastosowania określonej metody próbkowania, a jej wybór mógłby z kolei mieć duży wpływ na otrzymane wyniki (Veganzones i Séverin, 2018).

Dane finansowe wykorzystane w analizie dotyczyły lat 2015–2019 (rok przed złożeniem do sądu wniosku o ogłoszenie upadłości). Zbudowano bazę danych, na podstawie których obliczono wartości 27 wskaźników finansowych odnoszących się do podstawowych obszarów działalności firm, tj. rentowności, płynności, sprawności działania i struktury kapitałowej. Po wyeliminowaniu wskaźników silnie skorelowanych ostatecznie w badaniu uwzględniono 20 z nich (zestawienie 1).

[2] https://www.emis.com/pl.

Zestawienie 1. Wskaźniki finansowe wykorzystane w badaniu

Obszary	Wskaźnik finansowy	Formuła
Rentowność	rentowność aktywów (ROA) w % rentowność kapitału własnego (ROE) w % rentowność sprzedaży (ROS) w % marża operacyjna w %	zysk netto : aktywa ogółem zysk netto : kapitał własny zysk netto : przychody ze sprzedaży zysk operacyjny : przychody ze sprzedaży
Sprawność	rotacja zapasów w dniach rotacja należności w dniach rotacja aktywów w dniach rotacja zobowiązań w dniach rotacja kapitału obrotowego	zapasy : przychody ze sprzedaży · 365 należności : przychody ze sprzedaży · 365 aktywa : przychody ze sprzedaży · 365 zobowiązania : przychody ze sprzedaży · 365 przychody ze sprzedaży : (aktywa obrotowe – zobowiązania krótkoterminowe)
Płynność	bieżąca płynność szybka płynność płynność gotówkowa	aktywa obrotowe : zobowiązania krótkoterminowe (aktywa obrotowe – zapasy) : zobowiązania krótkoterminowe (środki pieniężne i ich ekwiwalenty + szybko zbywalne papiery wartościowe) : zobowiązania krótkoterminowe
Struktura finansowa	ogólne zadłużenie w % wskaźniki: długu do kapitału w % gotówki do aktywów w % należności do aktywów w % zapasów do aktywów w % rzeczowych aktywów trwałych do aktywów w % zobowiązań długoterminowych do zobowiązań ogółem w % kapitału do aktywów w %	zobowiązania : aktywa dług : kapitał własny gotówka : aktywa należności : aktywa zapasy : aktywa rzeczowe aktywa trwałe : aktywa zobowiązania długoterminowe : zobowiązania kapitał własny : aktywa

Źródło: opracowanie własne.

Do konstruowania modeli prognozowania upadłości przedsiębiorstw wykorzystano sześć następujących metod doboru zmiennych:

- metody oparte na filtrach (ang. *filter methods*), wykorzystujące entropię (metoda oznaczana dalej jako „entropia"), statystykę Manna-Whitneya („Whitney") i algorytm ReliefF („relief")[3];
- metody oparte na selekcji modeli (ang. *wrappers*) – krokowe („postępująca" i „wsteczna"), bazujące na funkcji dyskryminacyjnej i kryterium poprawy trafności klasyfikacji;
- metody będące integralną częścią algorytmu uczącego (ang. *embedded methods*), bazujące na indeksie Giniego.

Modele prognozowania upadłości przedsiębiorstw zostały skonstruowane za pomocą sześciu metod klasyfikacyjnych, tj.: metody k-najbliższych sąsiadów (oznacza-

[3] Wybór sześciu najlepszych zmiennych pod względem danego kryterium.

nej dalej jako „KNN"), metody wektorów nośnych („SVM"), sieci neuronowych („sieci"), lasu losowego („las"), algorytmu baggingowego („bagging") i metody wzmacniania gradientowego („XGBoost"). W zestawieniu 2 przedstawiono najważniejsze założenia przyjęte przy wykorzystaniu wymienionych metod. W badaniu nie uwzględniono optymalizacji podanych poniżej parametrów, ponieważ jego celem było porównanie trafności klasyfikacji modeli sektorowych i modelu ogólnego, a nie uzyskanie najlepszego możliwego modelu prognostycznego.

Zestawienie 2. Założenia przyjęte dla poszczególnych metod klasyfikacyjnych

Metody	Założenia
KNN	k optymalizowane za pomocą walidacji krzyżowej, maksymalna liczba najbliższych sąsiadów: 30
SVM	funkcja jądrowa: radialna
Sieci	jedna warstwa ukryta, liczba neuronów: średnia z liczby neuronów warstwy wejściowej i wyjściowej
Las	liczba drzew: 500, liczba zmiennych: 5
Bagging	liczba prób bootstrapowych: 500
XGBoost	liczba iteracji: 500, $\eta = 0{,}001$

Źródło: opracowanie własne.

Analiza została w całości przeprowadzona przy użyciu własnej aplikacji internetowej z graficznym interfejsem użytkownika, wykonanej w środowisku R. Może ona być stosowana w przyszłości do rozwiązania dowolnego problemu klasyfikacji obiektów wielowymiarowych.

4. Wyniki badania

W pierwszym kroku badania zweryfikowano, czy badane przedsiębiorstwa – bankruci i podmioty zdrowe – różnią się pod względem kondycji finansowej, opisanej za pomocą przedstawionych wcześniej wskaźników finansowych. Ze względu na to, że wartości wskaźników nie charakteryzowały się rozkładem normalnym (o czym świadczyły wyniki przeprowadzonych czterech testów normalności rozkładu), wykorzystano w tym celu nieparametryczny test Manna-Whitneya. Wyniki dla całej próby badawczej przedstawiono w tabl. 1.

Tabl. 1. Analiza porównawcza wybranych wskaźników finansowych dla badanych grup

Wskaźniki finansowe	Symbol	Mediana		Statystyka Z^a	Wartość p
		bankruci	podmioty zdrowe		
Rentowność aktywów (ROA) w %	V1	–27,90	–4,57	–12,39	0,00
Rentowność kapitału własnego (ROE) w %	V2	20,66	18,65	4,84	0,00
Rentowność sprzedaży (ROS) w %	V3	–17,52	–2,31	–10,48	0,00
Marża operacyjna w %	V4	–14,66	–1,16	–10,18	0,00
Rotacja zapasów w dniach	V5	8,16	48,51	–7,13	0,00
Rotacja należności w dniach	V6	72,43	33,56	–0,15	0,88
Rotacja aktywów w dniach	V7	242,92	181,45	–1,97	0,05
Rotacja zobowiązań w dniach	V8	15,12	9,05	1,12	0,26
Rotacja kapitału obrotowego	V9	–0,41	3,87	–7,99	0,00
Bieżąca płynność	V10	0,60	1,41	–14,03	0,00
Szybka płynność	V11	0,51	0,73	–9,14	0,00
Płynność gotówkowa	V12	0,01	0,08	–15,73	0,00
Ogólne zadłużenie w %	V13	16,26	29,75	–1,50	0,13
Wskaźniki: długu do kapitału w %	V14	0,00	16,61	–7,39	0,00
gotówki do aktywów w %	V15	1,36	4,97	–10,43	0,00
należności do aktywów w %	V16	26,49	16,73	2,17	0,03
zapasów do aktywów w %	V17	3,89	29,84	–7,30	0,00
rzeczowych aktywów trwałych do aktywów w %	V18	3,45	15,07	–4,83	0,00
zobowiązań długoterminowych do zobowiązań ogółem w %	V19	1,96	21,54	–12.20	0,00
kapitału do aktywów w %	V20	–24,57	18,88	–13,21	0,00

a Wartość statystyki testowej dla testu U Manna-Whitneya.

Źródło: opracowanie własne na podstawie danych z bazy EMIS Professional.

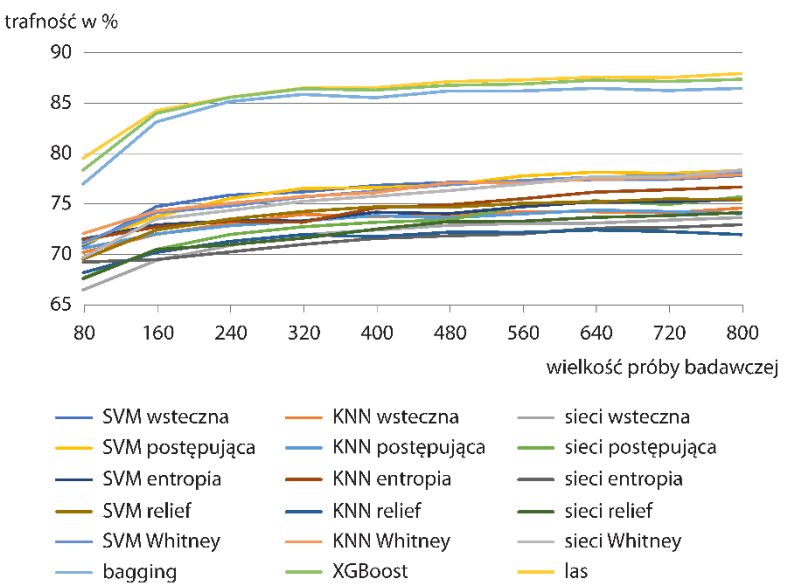
Na podstawie uzyskanych wyników można zauważyć, że badane przedsiębiorstwa różnią się istotnie (przy poziomie istotności równym 0,05) pod względem 17 z 20 badanych wskaźników. Podmioty zdrowe cechuje przede wszystkim wyższa rentowność i płynność działania niż bankrutów. Można uznać, że wybrane zmienne charakteryzują się odpowiednią siłą dyskryminacyjną i zasadne jest ich uwzględnienie w konstrukcji modeli prognozowania upadłości przedsiębiorstw.

Jak wspomniano, modele zostały skonstruowane za pomocą sześciu metod klasyfikacyjnych. W przypadku metod k-najbliższych sąsiadów, wektorów nośnych i sieci neuronowych zmienne zostały dobrane z wykorzystaniem wcześniej wspomnianych metod doboru opartych na filtrach (entropia, Whitney i relief) oraz na selekcji zmiennych do modeli (postępująca i wsteczna). W ten sposób uzyskano wyniki dla 15 kombinacji metody klasyfikacji i metody doboru zmiennych. W przypadku lasu, baggingu i XGBoostu wykorzystano jedną metodę doboru zmiennych, bazującą na indeksie Giniego. W ten sposób uzyskano łącznie 18 wariantów wyników. Zastosowane podejście służyło sprawdzeniu odporności rezultatów badania na metodykę badawczą. Konstrukcja modelu klasyfikacyjnego wymagała podziału dostępnej próby badawczej na próby uczącą i testową. We wszystkich przeprowadzonych analizach ustalono, że obiekty są losowane 500-krotnie z zachowaniem proporcji 80 : 20 na korzyść próby uczącej.

W celu zweryfikowania zależności pomiędzy wielkością próby badawczej a trafnością klasyfikacji modeli skonstruowanych za pomocą przedstawionej metodyki wykorzystano

całą dostępną próbę 800 przedsiębiorstw, bez uwzględniania ich przynależności sektorowej. Modele te zostały zbudowane 500 razy na podstawie losowo dobieranej próby o różnej liczebności. Na wykr. 1 przedstawiono uśrednione (dla 500 iteracji) wskaźniki globalnej trafności, przedstawiające udział poprawnie zaklasyfikowanych przedsiębiorstw – bankrutów i podmiotów zdrowych – we wszystkich obiektach poddanych klasyfikacji, otrzymane dla 500 prób testowych. Na podstawie uzyskanych wyników można stwierdzić, że w miarę wzrostu wielkości próby badawczej trafność globalna dla każdej zastosowanej metody się poprawia. Tempo poprawy jest zróżnicowane – początkowo jest znacznie wyższe niż dla większych prób badawczych. Najwyższą trafnością klasyfikacji charakteryzują się trzy metody bazujące na drzewach klasyfikacyjnych: las, bagging i XGBoost. Rezultaty przeprowadzonej weryfikacji pozwalają wyciągnąć wniosek, że jakość prognostyczna modelu klasyfikacyjnego jest uzależniona od wielkości próby badawczej. Ma to szczególnie duże znaczenie w przypadku małej liczby obserwacji.

Wykr. 1. Globalna trafność klasyfikacji uzyskana dla różnych wielkości prób i metodyk badawczych

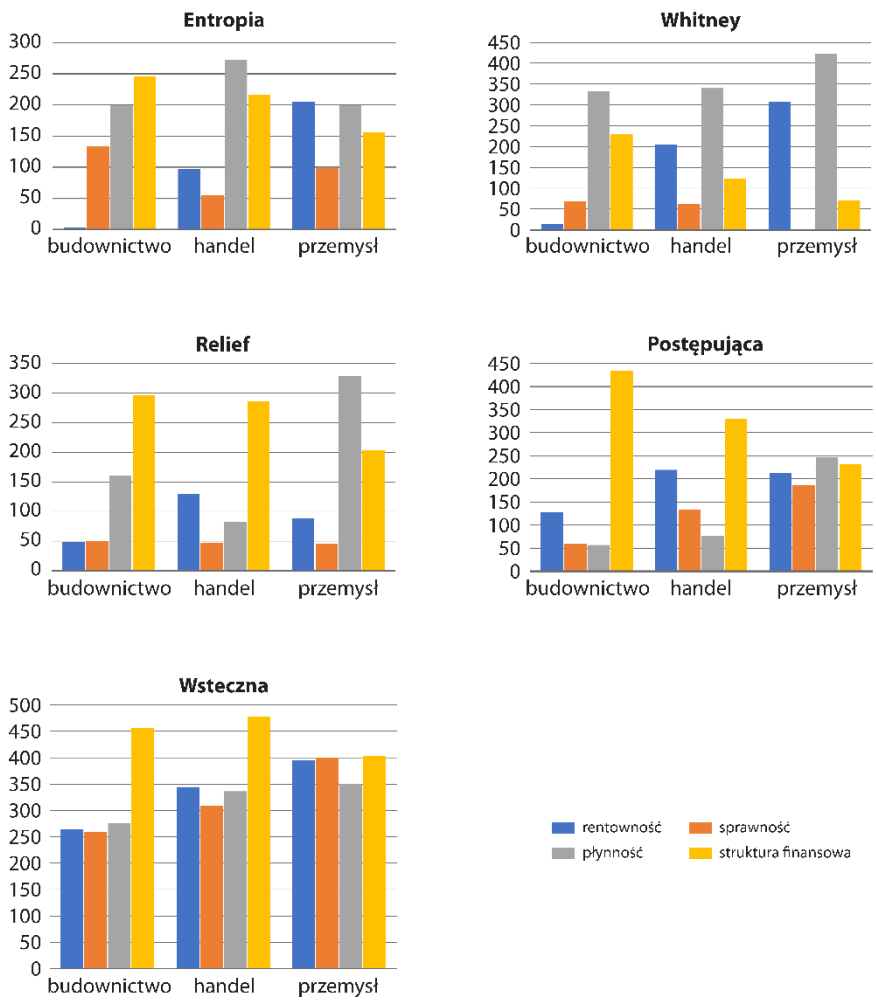


Źródło: opracowanie własne na podstawie danych z bazy EMIS Professional.

Opierając się na powyższym wniosku, stwierdzono, że porównywanie modeli sektorowych i modelu ogólnego (nieuwzględniającego specyfiki sektorowej) jest zasadne tylko wówczas, gdy modele są konstruowane na podstawie prób badawczych tej samej wielkości. Ze względu na dostępność danych postanowiono, że liczba przedsiębiorstw w próbie będzie wynosić 200. I tak 500 razy losowano 200 przedsiębiorstw z poszczególnych sektorów gospodarki do prób uczącej i testowej w celu skonstruowania modeli sektorowych oraz analogicznie 500 razy losowano 200 obiektów z całej

dostępnej próby 800 przedsiębiorstw w celu skonstruowania modelu ogólnego. Dla każdego losowania, na podstawie próby uczącej, dobrano wskaźniki finansowe z wykorzystaniem wcześniej opisanych metod. Na podstawie tych zmiennych konstruowano modele prognostyczne, których jakość została sprawdzona w próbach testowych. Na wykr. 2 przedstawiono średnią częstość wyboru wskaźników finansowych dla poszczególnych obszarów funkcjonowania firm, z wykorzystaniem przedstawionych wcześniej pięciu metod doboru zmiennych.

Wykr. 2. Przeciętna częstość wykorzystania wskaźników finansowych dotyczących różnych obszarów działania przedsiębiorstwa w konstruowaniu modeli klasyfikacyjnych dla wybranych sektorów gospodarki i metod doboru zmiennych



Źródło: opracowanie własne na podstawie danych z bazy EMIS Professional.

Uzyskane wyniki pozwalają na stwierdzenie, że modele konstruowane dla poszczególnych sektorów gospodarki różnią się średnią częstością wykorzystania wskaźników finansowych. Wskaźniki odnoszące się do rentowności – bez względu na metodę doboru zmiennych – były przeciętnie rzadziej wybierane do modelu dla budownictwa niż dla pozostałych sektorów gospodarki. Z kolei zmienne opisujące strukturę finansową pojawiały się w przypadku tego sektora częściej niż w przypadku przetwórstwa przemysłowego i handlu. Dla przetwórstwa przemysłowego charakterystyczne jest zaś to, że rzadziej niż w przypadku dwóch pozostałych sektorów do modelu wybierane są wskaźniki dotyczące struktury finansowej.

W kolejnym kroku badania porównano przeciętną (dla 500 iteracji) trafność klasyfikacji uzyskaną dla modelu ogólnego i modeli sektorowych. Przeciętna wartość wskaźników globalnej trafności otrzymana dla modeli została przedstawiona w tabl. 2.

Tabl. 2. Przeciętna globalna trafność klasyfikacji uzyskana dla modelu ogólnego i modeli sektorowych

Metody	Trafność w %		Statystyka Z^a	Wartość p
	model ogólny	modele sektorowe		
Las	85,47	85,30	1,17	0,24
Bagging	84,59	83,53	3,87	0,00
XGBoost	83,86	84,56	–2,08	0,04
SVM wsteczna	74,74	75,84	–3,08	0,00
KNN wsteczna	72,27	73,63	–4,28	0,00
Sieci wsteczna	69,80	71,78	–4,76	0,00
SVM postępująca	71,66	70,50	–3,29	0,00
KNN postępująca	71,84	73,40	–4,22	0,00
Sieci postępująca	70,04	71,57	–3,62	0,00
SVM entropia	73,34	74,51	–2,46	0,01
KNN entropia	73,00	74,51	–3,30	0,00
Sieci entropia	69,91	71,58	–4,32	0,00
SVM relief	73,38	74,57	–2,91	0,00
KNN relief	70,70	72,90	–5,67	0,00
Sieci relief	69,53	71,33	–3,48	0,00
SVM Whitney	74,07	75,47	–3,83	0,00
KNN Whitney	74,41	74,51	–0,24	0,81
Sieci Whitney	73,61	74,47	–2,51	0,01

a Jak przy tabl. 1.

Źródło: opracowanie własne na podstawie danych z bazy EMIS Professional.

W przeważającej większości modeli sektorowych globalna trafność klasyfikacji jest wyższa niż dla modeli ogólnych (14 razy różnice te okazały się statystycznie istotne, przy poziomie istotności 0,05). Wyłącznie w dwóch przypadkach (bagging i SVM postępująca) modele ogólne charakteryzowały się przeciętnie wyższą, statystycznie istotną, globalną trafnością klasyfikacji.

W celu dokładniejszej analizy uzyskanych wyników zweryfikowano, jak sprawnie analizowane modele działają w przypadku podmiotów zdrowych i bankrutów. Dane dotyczące trafności klasyfikacji przedstawiono w tabl. 3.

Tabl. 3. Przeciętna trafność klasyfikacji uzyskana dla modelu ogólnego i modeli sektorowych dla podmiotów zdrowych i bankrutów

Metody	Trafność w %		Statystyka Z[a]	Wartość p
	model ogólny	modele sektorowe		
Podmioty zdrowe				
Las	85,30	84,03	4,46	0,00
Bagging	86,93	84,83	6,19	0,00
XGBoost	86,45	86,52	1,83	0,07
SVM wsteczna	76,50	78,82	–4,02	0,00
KNN wsteczna	80,61	83,41	–7,46	0,00
Sieci wsteczna	72,29	73,38	–0,81	0,35
SVM postępująca	73,58	75,38	–3,17	0,03
KNN postępująca	76,03	77,32	–6,42	0,18
Sieci postępująca	73,90	74,88	0,43	0,95
SVM entropia	73,29	74,62	–0,88	0,38
KNN entropia	78,74	78,46	1,65	0,10
Sieci entropia	72,77	74,94	–2,83	0,00
SVM relief	74,99	75,18	–0,32	0,75
KNN relief	76,14	77,70	–2,52	0,01
Sieci relief	73,43	74,80	–0,08	0,94
SVM Whitney	75,91	75,22	1,73	0,08
KNN Whitney	79,72	78,38	3,38	0,00
Sieci Whitney	77,74	78,26	0,09	0,93
Bankruci				
Las	85,67	86,60	–1,25	0,21
Bagging	82,24	82,22	0,86	0,39
XGBoost	81,28	82,59	–2,39	0,02
SVM wsteczna	72,98	72,85	0,20	0,84
KNN wsteczna	63,92	63,84	0,23	0,82
Sieci wsteczna	67,30	70,18	–3,02	0,00
SVM postępująca	69,73	65,62	5,57	0,00
KNN postępująca	67,65	69,48	–2,81	0,01
Sieci postępująca	66,18	68,27	–1,45	0,15
SVM entropia	73,39	74,39	–1,60	0,11
KNN entropia	67,25	70,56	–6,14	0,00
Sieci entropia	67,05	68,21	–0,98	0,32
SVM relief	71,76	73,96	–3,54	0,00
KNN relief	65,25	68,11	–3,98	0,00
Sieci relief	65,64	67,86	–1,62	0,11
SVM Whitney	72,22	75,71	–5,77	0,00
KNN Whitney	69,10	70,63	–2,15	0,03
Sieci Whitney	69,48	70,68	–1,26	0,21

a Jak przy tabl. 1.

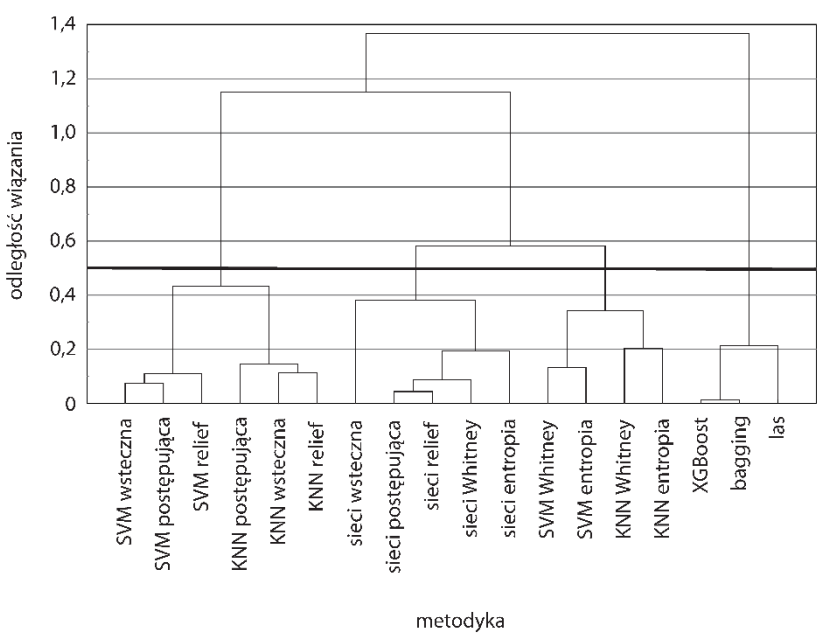
Źródło: opracowanie własne na podstawie danych z bazy EMIS Professional.

Można zaobserwować, że modele sektorowe, tak jak w przypadku globalnej trafności klasyfikacji, mają zdecydowaną przewagę nad modelami nieuwzględniającymi specyfiki działalności, jednak różnice między przeciętną trafnością klasyfikacji modeli ogólnych i sektorowych są dużo częściej nieistotne statystycznie. W przypadku podmiotów zdrowych modele sektorowe okazały się pięciokrotnie sprawniejsze pod

względem istotności statystycznej niż modele ogólne, a w trzech przypadkach miała miejsce sytuacja odwrotna. W przypadku bankrutów przewaga modeli sektorowych nad modelami ogólnymi była zdecydowanie wyraźniejsza – w ośmiu przypadkach uzyskały one wyższe wskaźniki trafności klasyfikacji, a różnice te były statystycznie istotne. Tylko raz modele ogólne osiągnęły istotną statystycznie wyższą trafność klasyfikacji. Rezultaty te są zbieżne z wynikami przedstawionymi przez innych autorów, przy czym w omawianym badaniu uwzględniono liczebność próby badawczej, różną metodykę badawczą oraz losowe próby uczące i testowe.

W ostatnim kroku analizy zbadano podobieństwo działania poszczególnych metod badawczych. W tym celu wykorzystano wcześniejsze wyniki klasyfikacji dla 500-krotnego losowania i modeli skonstruowanych dla wszystkich 800 badanych przedsiębiorstw. Dysponując informacją o tym, jak często losowane obiekty zostały zaklasyfikowane poprawnie lub błędnie, obliczono macierz korelacji pomiędzy tymi częstościami dla zastosowanych metod badawczych. Następnie przekształcono macierz korelacji w macierz odległości, które obliczono poprzez odjęcie wartości danego współczynnika korelacji od 1. Na podstawie tak uzyskanej macierzy i jednej z aglomeracyjnych metod analizy skupień – metody Warda – skonstruowano dendrogram (wykr. 3).

Wykr. 3. Podobieństwo działania poszczególnych metod badawczych



Źródło: opracowanie własne na podstawie danych z bazy EMIS Professional.

Na podstawie wykr. 3 można stwierdzić, że istnieją grupy metod bardzo podobnych do siebie pod względem sposobu działania. Przyjmując jako punkt podziału wiązanie o długości 0,5 (linia prosta na wykresie), można wyróżnić cztery skupienia. Tworzą je dwa skupienia obejmujące modele zbudowane z wykorzystaniem metody wektorów nośnych i *k*-najbliższych sąsiadów, skupienie obejmujące modele sieci neuronowych oraz skupienie modeli bazujących na drzewach klasyfikacyjnych. Okazuje się więc, że zastosowana metoda doboru zmiennych do modelu nie ma w tym przypadku większego znaczenia.

5. Podsumowanie

Przeprowadzone badanie dotyczyło 400 przedsiębiorstw, wobec których wszczęto postępowanie upadłościowe w Polsce w latach 2016–2020, oraz 400 przedsiębiorstw w dobrej kondycji finansowej. Zasadniczym celem analizy było uzyskanie odpowiedzi na pytanie, czy konstruowanie modeli uwzględniających kryterium przynależności sektorowej jest zasadne. Aby to sprawdzić, wykorzystano sześć metod doboru zmiennych, sześć metod klasyfikacyjnych oraz 500 losowych prób uczących i testowych.

Z badania wynika, że prognozując upadłość przedsiębiorstw, powinno się uwzględniać rodzaj prowadzonej działalności gospodarczej, ponieważ w poszczególnych sektorach gospodarki istnieją odmienne determinanty upadłości. Modele skonstruowane dla grupy przedsiębiorstw prowadzących jednorodną działalność gospodarczą pozwalają uzyskać przeciętnie wyższe wartości wskaźników trafności klasyfikacji. Biorąc pod uwagę otrzymane rezultaty, można także stwierdzić, że trafność klasyfikacji modelu prognostycznego jest uzależniona od wielkości próby badawczej, a modele konstruowane z wykorzystaniem tych samych metod prognostycznych klasyfikują badane obiekty w podobny sposób.

Bibliografia

Altman, E. I. (1983). *Corporate financial distress: A complete guide to predicting, avoiding, and dealing with bankruptcy*. John Wiley & Sons.

Appenzeller, D., Nowara, W. (2003). Modele prognozowania upadłości banków komercyjnych w Polsce. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, (1001), 13–22.

Bai, Q., Tian, S. (2020). Innovate or die: Corporate innovation and bankruptcy forecasts. *Journal of Empirical Finance*, *59*, 88–108. https://doi.org/10.1016/j.jempfin.2020.09.002.

Becerra-Vicario, R., Alaminos, D., Aranda, E., Fernández-Gámez, M. A. (2020). Deep Recurrent Convolutional Neural Network for Bankruptcy Prediction: A Case of the Restaurant Industry. *Sustainability*, *12*(12), 1–15. https://doi.org/10.3390/su12125180.

Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C., Popp, J. (2013). Sample size planning for classification models. *Analytica Chimica Acta*, *760*, 25–33. https://doi.org/10.1016/j.aca.2012.11.007.

Bellovary, J., Giacomino, D., Akers, M. (2007). A Review of Bankruptcy Prediction Studies: 1930 to Present. *Journal of Financial Education*, *33*(4), 1–42. https://www.jstor.org/stable/41948574.

Boratyńska, K., Grzegorzewska, E. (2018). Bankruptcy prediction in the agribusiness sector: Lessons from quantitative and qualitative approaches. *Journal of Business Research*, *89*, 175–181. https://doi.org/10.1016/j.jbusres.2018.01.028.

Brożyna, J., Mendel, G., Pisula, T. (2016). Statistical methods of the bankruptcy prediction in the logistic sector in Poland and Slovakia. *Transformations in Business & Economics*, *15*(1), 93–114.

Campbell, J. Y., Hilscher, J., Szilagyi, J. (2008). In Search of Distress Risk. *The Journal of Finance*, *63*(6), 2899–2939. https://doi.org/10.1111/j.1540-6261.2008.01416.x.

Chan, C. Y., Chou, D. W., Lin, J. R., Liu, F. Y. (2016). The role of corporate governance in forecasting bankruptcy: Pre- and post-SOX enactment. *The North American Journal of Economics and Finance*, *35*, 166–188. https://doi.org/10.1016/j.najef.2015.10.008.

Chava, S., Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, *8*(4), 537–569. https://doi.org/10.1093/rof/8.4.537.

Chen, Y., Zhang, L., Zhang, L. (2013). Financial Distress Prediction for Chinese Listed Manufacturing Companies. *Procedia Computer Science*, *17*, 678–686.

Dudycz, T., Skoczylas, W. (2023). Wskaźniki finansowe przedsiębiorstw według działów (sektorów) za 2021 r. *Rachunkowość*, (5), 61–88.

Gabbianelli, L. (2018). A territorial perspective of SME's default prediction models. *Studies in Economics and Finance*, *35*(4), 542–563. https://doi.org/10.1108/SEF-08-2016-0207.

Gajdka, J., Stos, D. (1996). Wykorzystanie analizy dyskryminacyjnej do badania podatności przedsiębiorstwa na bankructwo. W: J. Duraj (red.), *Przedsiębiorstwo na rynku kapitałowym (*s. 138–148). Polskie Wydawnictwo Ekonomiczne.

Gill de Albornoz, B., Giner, B. (2013). Predicción del fracaso empresarial en los sectores construcción e inmobiliario: Modelos generales versus específicos. *Universia Business Review*, (39), 118–131.

Grzegorzewska, E. (2011). Zagrożenie upadłością a cykl życia przedsiębiorstw rolniczych. W: E. Mączyńska (red.), *Cykle życia i bankructwa przedsiębiorstw* (s. 281–301). Oficyna Wydawnicza SGH.

Grzegorzewska, E., Runowski, H. (2008). Zdolności prognostyczne polskich modeli dyskryminacyjnych w badaniu kondycji finansowej przedsiębiorstw rolniczych. *Roczniki Nauk Rolniczych, Seria G – Ekonomika Rolnictwa*, *95*(3/4), 83–90. https://sj.wne.sggw.pl/pdf/RNR_2008_n3-4_s83.pdf.

Hadasik, D. (1998). *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*. Wydawnictwo Akademii Ekonomicznej w Poznaniu.

Heo, J., Yang, J. Y. (2014). AdaBoost based bankruptcy forecasting of Korean construction companies. *Applied Soft Computing*, *24*, 494–499. https://doi.org/10.1016/j.asoc.2014.08.009.

Herman, S. (2017). Industry specifics of joint-stock companies in Poland and their bankruptcy prediction. W: M. Papież, S. Śmiech (red.), *Proceedings of the 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena* (s. 93–102). Fundacja Uniwersytetu Ekonomicznego w Krakowie.

Hołda, A. (2001). Prognozowanie bankructwa jednostki w warunkach polskiej gospodarki z wykorzystaniem funkcji dyskryminacyjnej. *Rachunkowość*, (5), 306–310.

Hołda, A. (2006). *Zasada kontynuacji działalności i prognozowanie upadłości w polskich realiach gospodarczych*. Wydawnictwo Akademii Ekonomicznej w Krakowie.

Hołda, A. (2009). Wykorzystanie drzew decyzyjnych w prognozowaniu upadłości przedsiębiorstw w branży budowlanej. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, (796), 165–175.

Iwanowicz, T. (2018). Empiryczna weryfikacja hipotezy o przenośności modelu Altmana na warunki polskiej gospodarki oraz uniwersalności sektorowej modeli. *Zeszyty Teoretyczne Rachunkowości*, (96), 63–80. https://doi.org/10.5604/01.3001.0011.6170.

Jabeur, B. S., Stef, N., Carmona, P. (2022). Bankruptcy Prediction using the XGBoost Algorithm and Variable Importance Feature Engineering. *Computational Economics*, *61*, 715–741. https://doi.org/10.1007/s10614-021-10227-1.

Jagiełło, R. (2013). *Analiza dyskryminacyjna i regresja logistyczna w procesie oceny zdolności kredytowej przedsiębiorstw*. Narodowy Bank Polski. https://static.nbp.pl/publikacje/materialy-i-studia/ms286.pdf.

du Jardin, P. (2018). Failure pattern-based ensembles applied to bankruptcy forecasting. *Decision Support Systems*, *107*, 64–77. https://doi.org/10.1016/j.dss.2018.01.003.

Juszczyk, S. (2010). Prognozowanie upadłości przedsiębiorstw. *Ekonomista*, (5), 701–728.

Juszczyk, S., Balina, R. (2014). Prognozowanie zagrożenia bankructwem przedsiębiorstw w wybranych branżach. *Ekonomista*, (1), 67–95. https://ekonomista.pte.pl/pdf-155672-82501?filename=Prognozowanie%20zagrozenia.pdf.

Karbownik, L. (2013). The Use of Accrual-Based and Cash-Based Approach in Evaluating the Operational Financial Threat of Enterprises from the TSL Sector – Example of Application of the Discriminant Analysis. *Quantitative Methods in Economics. Metody Ilościowe w Badaniach Ekonomicznych*, *14*(1), 190–201. https://qme.sggw.edu.pl/article/view/3588.

Korol, T. (2005). Wykorzystanie sieci jednokierunkowej wielowarstwowej oraz sieci rekurencyjnej w prognozowaniu upadłości przedsiębiorstw. *Materiały i Prace Instytutu Funkcjonowania Gospodarki Narodowej*, *93*, 169–181.

Korol, T., Prusak, B. (2005). *Upadłość przedsiębiorstw a wykorzystanie sztucznej inteligencji*. CeDeWu.

Król, K., Stefański, A. (2014). Metodyka budowy modelu prognozowania bankructwa na przykładzie sektora budowlanego. *Zeszyty Naukowe Wyższej Szkoły Bankowej we Wrocławiu*, (7), 159–184.

Laguillo, G., del Castillo, A., Fernández, M. Á., Becerra, R. (2019). Focused vs unfocused models for bankruptcy prediction: Empirical evidence for Spain. *Contaduría y Administración*, *64*(2), 1–22. https://doi.org/10.22201/fca.24488410e.2018.1488.

Liang, D., Lu, C. C., Tsai, C. F., Shih, G. A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, *252*(2), 561–572. https://doi.org/10.1016/j.ejor.2016.01.012.

Liang, D., Tsai, C. F., Wu, H. T. (2015). The effect of feature selection on financial distress prediction. *Knowledge-Based Systems*, *73*(1), 289–297. https://doi.org/10.1016/j.knosys.2014.10.010.

Mroczkowska, A., Rogowski, W. (2010). Światowe doświadczenia w zakresie tworzenia modeli prognozowania zagrożenia przedsiębiorstwa upadłością. *Studia i Prace Kolegium Zarządzania i Finansów*, (102), 42–69. https://econjournals.sgh.waw.pl/public/journals/5/archiwum_2018_2014/2010/ZN_102.pdf.

Noga, T., Adamowicz, K., Jakubowski, J. (2014). Metody dyskryminacyjne w ocenie sytuacji finansowej przedsiębiorstw sektora leśno-drzewnego. *Acta Scientiarum Polonorum. Silvarum Colendarum Ratio et Industria Lignaria*, *13*(1), 25–35.

Nyitrai, T., Virág, M. (2019). The effects of handling outliers on the performance of bankruptcy prediction models. *Socio-Economic Planning Sciences*, *67*, 34–42. https://doi.org/10.1016/j.seps.2018.08.004.

Pisula, T. (2010a). Prognozowanie zagrożenia bankructwem dla spółek giełdowych z sektora budowlanego. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (117), 351–360.

Pisula, T. (2010b). Prognozowanie zagrożenia upadłością dla polskich spółek giełdowych z sektora deweloperów z wykorzystaniem modeli strukturalnych. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Finanse, Rynki Finansowe, Ubezpieczenia*, (29), 109–122.

Pisula, T. (2017). Zastosowanie ensemble klasyfikatorów do oceny ryzyka upadłości przedsiębiorstw na przykładzie firm sektora produkcyjnego działających na Podkarpaciu. *Zarządzanie i Finanse*, *15*(3), 279–293.

Pociecha, J. (red.). (2014). *Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej*. Fundacja Uniwersytetu Ekonomicznego w Krakowie.

Pogodzińska, M., Sojak, S. (1995). Wykorzystanie analizy dyskryminacyjnej w przewidywaniu bankructwa przedsiębiorstw. *Acta Universitatis Nicolai Copernici. Nauki Humanistyczno-Społeczne. Ekonomia*, (25), 53–61.

Potoczna, M., Wiśniewska, O. (2013). Zastosowanie analizy dyskryminacyjnej oraz modelu logitowego do prognozowania upadłości polskich przedsiębiorstw. W: S. Wawak (red.), *Metody i techniki diagnostyczne w doskonaleniu organizacji* (s. 39–47). Mfiles.pl.

Prusak, B. (2005). *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*. Difin.

Ptak-Chmielewska, A. (2018). Bankruptcy Risk Models for Polish SMEs – Regional Approach. *Acta Universitatis Lodziensis. Folia Oeconomica*, *1*(333), 71–83. https://doi.org/10.18778/0208-6018.333.05.

Rajin, D., Milenković, D., Radojević, T. (2016). Bankruptcy prediction models in the Serbian agricultural sector. *Ekonomika Poljoprivrede. Economics of Agriculture*, *63*(1), 89–105. http://dx.doi.org/10.5937/ekoPolj1601089R.

Ravi Kumar, P., Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research*, *180*(1), 1–28. https://doi.org/10.1016/j.ejor.2006.08.043.

Rusiecki, K., Białek-Jaworska, A. (2015). Systemy wczesnego ostrzegania o zagrożeniu upadłością przedsiębiorstw z sektora budowlanego – porównanie analizy dyskryminacyjnej i modelu logitowego. *Ekonomia. Rynek, Gospodarka, Społeczeństwo*, *43*, 137–160. https://doi.org/10.17451/eko/43/2015/126.

Vabalas, A., Gowen, E., Poliakoff, E., Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLoS ONE*, *14*(11), 1–20. https://doi.org/10.1371/journal.pone.0224365.

Veganzones, D., Séverin, E. (2018). An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems*, *112*, 111–124. https://doi.org/10.1016/j.dss.2018.06.011.

Wysocki, F., Kozera, A. (2012). Wykorzystanie analizy dyskryminacyjnej w ocenie ryzyka upadłości przedsiębiorstw przemysłu mięsnego. *Journal of Agribusiness and Rural Development*, *4*(26), 167–182.

Zhang, C., Soda, P., Bi, J., Fan, G., Almpanidis, G., García, S., Ding, W. (2022). An empirical study on the joint impact of feature selection and data resampling on imbalance classification. *Applied Intelligence*, *53*, 5449–5461. https://doi.org/10.1007/s10489-022-03772-1.

Zhou, L. (2013). Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems*, *41*, 16–25. https://doi.org/10.1016/j.knosys.2012.12.007.

Zielińska-Sitkiewicz, M. (2016). Zastosowanie metod wielowymiarowej analizy dyskryminacyjnej do prognozowania upadłości wybranych spółek sektora spożywczego. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie. Ekonomika i Organizacja Gospodarki Żywnościowej*, (113), 117–129. https://doi.org/10.22630/EIOGZ.2016.113.10.

Zoričák, M., Gnip, P., Drotár, P., Gazda, V. (2020). Bankruptcy prediction for small- and medium-sized companies using severely imbalanced datasets. *Economic Modelling*, *84*, 165–176. https://doi.org/10.1016/j.econmod.2019.04.003.

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(10), 59–74 https://doi.org/10.59139/ws.2024.10.4
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 6.09.2023, poprawiony/revised: 15.04.2024, zaakceptowany/accepted: 2.09.2024

Identifying age groups of Twitter users based on the specific characteristics of textposts[1]

Krzysztof Najman,[a] Kamila Migdał-Najman,[b] Katarzyna Raca,[c] Agata Majkowska[d]

Abstract. Textual data (textposts) account for a significant portion of all data posted on the Internet. One piece of information that researchers are seeking to obtain about the authors of textposts is their age, which is not always made public, yet important from the point of view of marketing, social and economic research. Language research shows that representatives of different age groups tend to use a distinct set of vocabulary and grammatical forms. Presumably, textpost formatting as well as the level of the correctness of the text itself may also differentiate user age groups. The aim of the research presented in this article is to use the elements typically eliminated from texts during text mining processes, such as emoticons, punctuation marks and words that are not content carriers (stopwords) to distinguish the age groups of the authors of Twitter (currently X) posts. The study analysed nearly 3 million tweets in English posted before July 2020. The research shows that distinguished textpost elements differentiate the age groups only to a small extent. The youngest users stood out the most due to their specific language characteristics in textposts.

Keywords: Twitter, text mining, user age
JEL: C38, C88, M30

Identyfikacja grup wieku użytkowników Twittera na podstawie charakterystyki wiadomości tekstowych

Streszczenie. Dane (wiadomości) tekstowe stanowią znaczną część wszystkich danych zamieszczanych w Internecie. Jedną z informacji, które badacze chcieliby uzyskać o autorach wiadomości tekstowych, jest ich wiek, ponieważ ma on duże znaczenie z perspektywy badań marketingowych, społecznych czy ekonomicznych. Nie zawsze jednak data urodzenia jest udostępniana publicznie. Z badań językowych wynika, że przedstawiciele różnych grup wieku posługują się odmiennym słownictwem i innymi formami gramatycznymi. Wydaje się, że mogą je różnicować również sposoby formatowania wiadomości tekstowych i poprawność zapisu tekstu. Celem badania

[1] Artykuł został opracowany na podstawie referatu wygłoszonego na XXXI Konferencji Naukowej Sekcji Klasyfikacji i Analizy Danych PTS, która odbyła się w dniach 7–9 września 2022 r. w Warszawie. / The article is based on a paper delivered at the XXXI Scientific Conference of the Section of Classification and Data Analysis of PTS, held on 7–9 September 2022 in Warsaw, Poland.

[a] Uniwersytet Gdański, Wydział Zarządzania, Polska / University of Gdańsk, Faculty of Management, Poland. ORCID: https://orcid.org/0000-0003-1673-2858. Autor korespondencyjny / Corresponding author, e-mail: krzysztof.najman@ug.edu.pl.
[b] Uniwersytet Gdański, Wydział Zarządzania, Polska / University of Gdańsk, Faculty of Management, Poland. ORCID: https://orcid.org/0000-0003-4106-2964. E-mail: kamila.migdal-najman@ug.edu.pl.
[c] Uniwersytet Gdański, Wydział Zarządzania, Polska / University of Gdańsk, Faculty of Management, Poland. ORCID: https://orcid.org/0000-0003-1760-1673. E-mail: katarzyna.raca@ug.edu.pl.
[d] Uniwersytet Gdański, Wydział Zarządzania, Polska / University of Gdańsk, Faculty of Management, Poland. ORCID: https://orcid.org/0000-0003-0336-3598. E-mail: agata.majkowska@ug.edu.pl.

omawianego w artykule jest wyodrębnienie grup wieku autorów wpisów na Twitterze (obecnie X) na podstawie elementów zwykle usuwanych z tekstów analizowanych metodami text mining, takich jak emotikony, znaki interpunkcyjne i słowa, które nie są nośnikami treści (ang. *stopwords*). Przeanalizowano prawie 3 mln tweetów w języku angielskim opublikowanych przed lipcem 2020 r. Badanie wykazało, że wyodrębnione cechy w niewielkim stopniu różnicują grupy wiekowe. Najbardziej specyficznym stylem pisania wiadomości wyróżniają się najmłodsi użytkownicy Internetu.
Słowa kluczowe: Twitter, text mining, wiek użytkowników

1. Introduction

Contemporarily, the term generation is commonly used both in colloquial speech as well as in scientific literature. Leopold von Ranke, in his work from 1514 entitled *Geschichten der romanischen und germanischen Völker* wrote that a whole new generation had emerged in Europe which was to completely change the face of history. The author emphasises that a generation has the power to create and reinforce new ideas, which that generation then owns (Wyka, 1977). Wilhelm Dilthey defines a generation as a peer group sharing common childhood and adolescence, for whom the stage of maturing occurs during the same period. A generation is made up of individuals who experience the same transformations and events, which took place during the period of their excitability. The source of the dissimilarity between one generation and another relates to their distinct experiences and perspectives on life (Wyka, 1939). Today, however, a generation is identified with a social group linked by a common bond and which acts together. It is characterised by specific patterns of behaviour as well as attitudes, views, values, aspirations, way of life or a different understanding of reality, visibly distinct from those of other groups. This definition draws attention to the temporal and qualitative aspects of the notion of generation (Sztompka, 2002).

These divergent generations must, nevertheless, communicate with one another, in an effort to reach a mutual and relatively satisfactory understanding. The development of mass culture and the Internet has become an important factor of change in the area of communication. These changes are also reflected in the language, as well as in the way new forms of communication are used. Textposts are one of such forms, acting as a kind of conveyor of the information on the changes taking place within the social sphere. It also serves as a source of information on the authors of the textposts. The linguistic and stylistic differences observed between various age groups in verbal and non-verbal communication characterise the current reality. However, changing the writing style of a generation is a long process. These changes appear to occur most rapidly among younger people. Paradoxically, this group, jointly with the civilisational achievements associated with the electronic reality, have caused these transformations to accelerate. The behaviour of young people in terms of communication also reflects certain inevitable global processes which may to some extent become a norm in the way people communicate. If this is the case, can we speak of a generational differentiation of language, possibly causing difficulties in intergenerational communication?

The tendencies observed in contemporary applied language, especially promoted among young people, include: the economisation of the language and the widespread pursuit of abbreviations and acronyms, formal and syntactic minimisation, the use of pictorial writing, the periodical emergence of linguistic fashions, the use of various types of borrowings from other languages, the mixing or unification of functional styles in communication, the hybridisation of genres and styles, the introduction of emotional communication, including vulgarisms, as well as emphasis on message lucidity (Jakobson, 1960). Communication has been taking on an incognito form, which encourages the authors of messages to freely express themselves, affecting the form, style and words used. Various features of language play, involving the use of message graphisation (combination of image and word), as well as such extra-linguistic elements as emoticons, photoblogs and graphics have become common in text messages and posts. This promotes the abandonment of conventional politeness formulas and phrases used towards the addressee, failure to follow the rules of correct spelling, including occurrence of spelling errors, as well as improper use of punctuation or diacritical marks (Goban-Klas, 2005).

The above phenomenon appears particularly intense in electronic or online communication. Text messages, mobile application notifications or even short messages sent via instant messaging platforms force abbreviated forms. The initial technical conditions restricted the length of single messages to a small number of characters. This enforced brevity of communication has popularised the use of emoticons, emojis, various language abbreviations and the abandonment of politeness forms. Although eventually technical limitations have been overcome, the abbreviated communication remained. A very good example of this process is Twitter (now transformed to X), one of the most widely used communicators, where several hundred million textposts are published every day. Its launch in 2006, along with the emergence and spread of smartphones which began in the year 2000, has significantly affected the direction and dynamics of the simplifications in communication (Rodrigues et al., 2018). Linguistic research focuses on identifying the differences in communication patterns among people of different ages. Various studies seek to examine the communication characteristics of those whose age and other demographic or social features were known. The attempt to identify the communicators' age based solely on the messages they post is of no less interest to researchers. When analysing microblogging content, relatively often researchers are faced with the problem of identifying the age of the author of a posted text. Knowing the author's age, at least to the extent where classifying him or her to a given age group is possible, is very useful from the point of view of e.g. marketing research, which provides information on the differences in purchasing preferences. In many cases, however, the authors of the content posted on the Internet do not specify their age, which significantly limits the profiling possibilities.

In 2021, Majkowska et al. determined that Twitter users from various age groups use different vocabulary and in 2022, these authors demonstrated that the graphic

signs commonly used in Twitter posts, such as emoticons and emojis, are also characteristic of specific age groups. The above-mentioned research was carried out on the basis of textual data and prepared in accordance with the principles of text mining analysis. This type of research utilises various measures involving the removal of certain text elements at the stage of preparing the dataset for analysis. The proceedings of the preliminary text preparation include:

- the conversion of uppercase letters to lowercase;
- the extraction of punctuation marks;
- the removal of words that do not carry content (such as pronouns);
- the elimination of all digits, hashtags, web addresses, other user mentions.

It is believed that such prepared data simplify the analysis, as the extracted elements do not usually carry any relevant information. While it is possible to agree with such a thesis, from the perspective of the understanding possibilities of the text, the elements commonly removed can still be important when the features of the message are analysed.

In light of the above, the aim of the research presented in this article is to use the elements typically eliminated from texts during text mining processes, such as emoticons, punctuation marks and words that are not content carriers (stopwords) to distinguish the age groups of the authors of Twitter posts.

2. Research method

In order to achieve the above-mentioned aim, relevant data was downloaded for analysis. The application programming interface (API) is a tool applied to download data from Twitter. The data can be accessed once an account is created and an access key obtained. A regular account ensures access only to a certain type of data and for a limited period. The Twitter API for Academic Research Access, however, enables downloading any type of tweets from an unrestricted time range. The study presented here covers more than 4 million tweets in English posted during the last two weeks of July 2020. The data includes posts from 14,224 users, with approximately 300 tweets per user. Each user was selected using the 'happy birthday to me' and 'I'm x years old' keywords, which allowed for the identification of age-specific Twitter users.

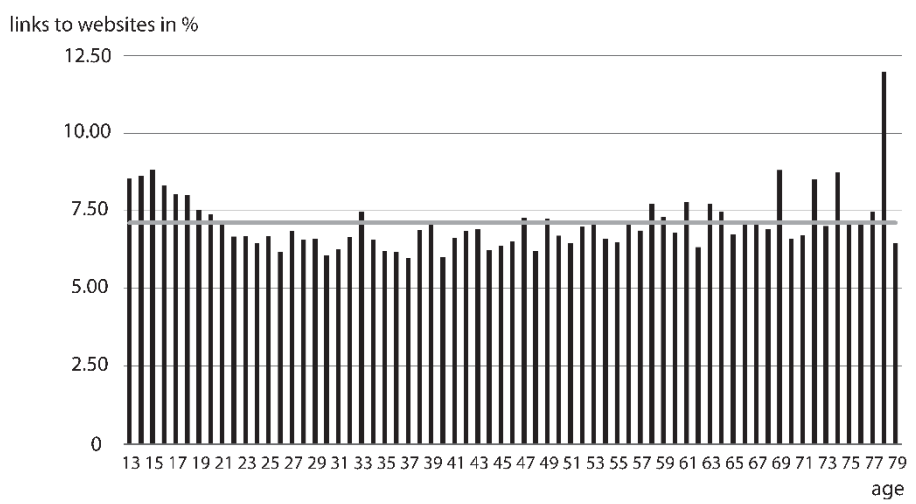
Prior to the analysis, the collected textual data required appropriate preparation. To achieve the set aim of the study, the pre-processing step of cleaning the data of certain elements was omitted. Punctuation marks, emoticons, stopwords, capital letters, digits, numbers written in words, hashtags, user mentions (@) and website links were extracted. In the next step, the number of a given elements' occurrence in each age group was counted.

The smiley face ':)' (7,631 occurrences) and the sad face ':(' (6,293 occurrences) accounted for almost half of all the emoticons posted by the studied Twitter users. The next position in the ranking was taken by the cat-face emoticon ':3', i.e. an imitation of

a cute facial expression (2,369 occurrences), 'XD' , expressing laughter (2,222 occurrences), and the surprised face 'oO' (2,132 occurrences). The most common punctuation marks identified in the analysed textposts were the period '.' (1,959,630 occurrences), the '@' symbol (1,831,703 occurrences) and the slash '/' (1,469,696 occurrences). The colon ':' (560,023 occurrences) and the comma ',' (555,667 occurrences) were the next most commonly used characters, but their number was about 3.5 times lower than that of the period. Twitter users are more likely to use their posts to refer to other users or share websites rather than tag topics or phrases, as shown by the fact that the hashtag symbol '#' was not in the top five most used punctuation marks. The preposition 'a' (9,474,296 occurrences) and the personal pronoun 'I' (7,505,693 occurrences) were the stopwords most frequently used by the studied Twitter users. This indicates most Twitter users' inclination to write about themselves or express their opinions. The preposition 'in' (1,854,896 occurrences), the personal pronoun 'he' (1,659,302 occurrences) and the preposition 'an' (1,591,886 occurrences) followed in the ranking. Other personal pronouns, such as 'she', 'we' and 'they', failed to make the top ten most frequently used stopwords.

On average, capital letters account for 7.11% of all characters used in Twitter users' posts. With regards to the occurrence of capital letters, the youngest and the oldest users stand out from the rest. This does not, however, mean that these users always begin writing their utterances with a capital letter. These statistics can also indicate the users' wish to emphasise the messages conveyed or the wording used through full capital-letter sentences or words. 37-year-olds are the least likely to use capital letters although this is not a strong disproportion in relation to the average level (see Figure 1).

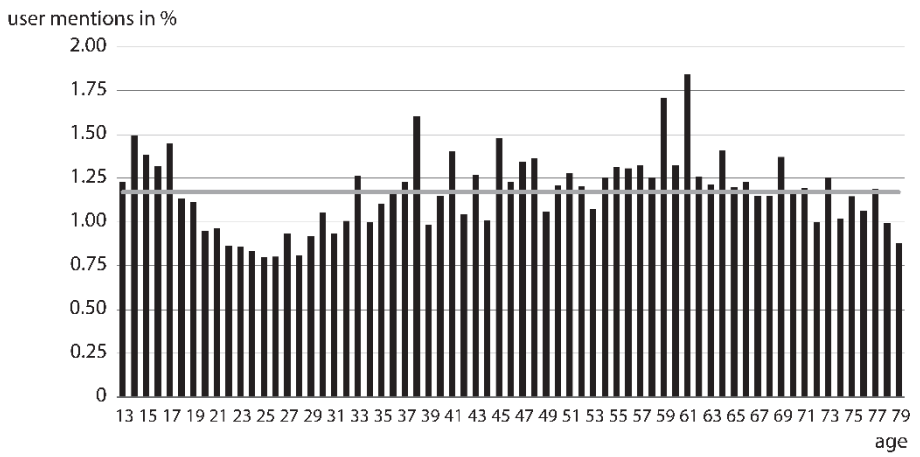
Figure 1. Percentage of capital letters used in tweets, by age



Source: authors' work based on data downloaded from Twitter.

References to other Twitter users account for 1.17% of all characters in a single textpost. The individual age groups do not differ significantly in the use of such references, with the exception of 61-year-olds, 59-year-olds and 38-year-olds. Among the representatives of 18- to 25-year-olds, a downward trend is observed in the share of this feature. Starting with the 26-year-olds onward, the trend shifts to an increasing one, so much so that 33-year-olds exceed in the average share of this trait (see Figure 2).

Figure 2. Percentage of user mentions in tweets, by age

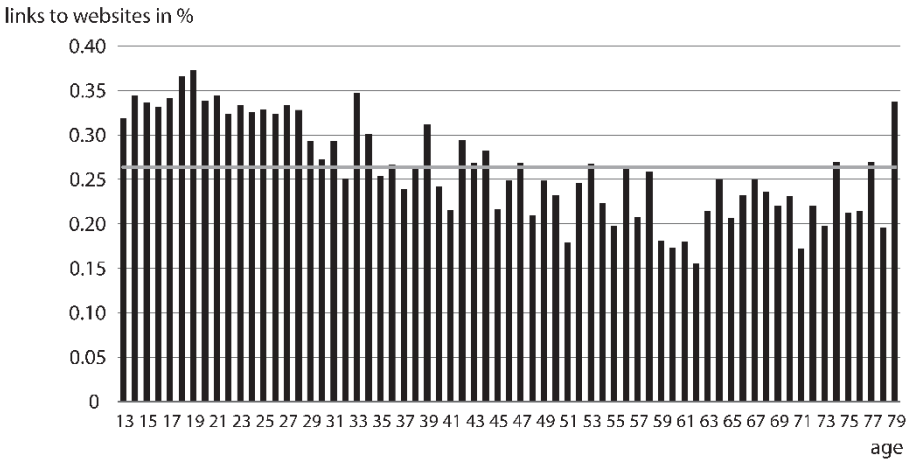


Source: authors' work based on data downloaded from Twitter.

Individual age groups of Twitter users do not differ significantly in terms of website link incidence. On average, web links account for 0.26% of all characters in a single textpost. 13- to 31-year-olds stand out in terms of website link use, as this proportion remains above the average level. This element is least frequently used in tweets by 62-year-olds (see Figure 3).

With regards to numerals incidence in textposts, the individual age groups do not differ significantly. Nevertheless, on average, 13- to 19-year-olds are distinguished by a lower proportion of numbers in word notation and a higher proportion of numbers written in digits. Moderate age variation is also observed with regards to the use of hashtags in tweets. On average, hashtags account for 0.13% in a single textpost. The 39-, 42-, 53- and 69-year-olds stand out in terms of hashtag use, while the 73-year-olds are the least likely to use such tagging.

Figure 3. Percentage use of website links in tweets, by age



Source: authors' work based on data downloaded from Twitter.

3. The essence and purposes of data classification

One of the oldest and most important human abilities is that of discerning and recognising different types of objects, including associating and combining these objects into classes. The ability to distinguish similar objects and group them is one such skill, used on a daily basis for object recognition or problem solving. This competence is natural for humans and has nowadays become the basis for the development of various fields of knowledge which provide a wide array of terminology and definitions. Carolus Linnaeus was right by stating that classification should facilitate object-to-group identification. According to John A. Hartigan, classification is a way of considering and perceiving things rather than the study of things as such (Arabie et al., 1996). Various definitions of the term classification have been propounded in the literature on the subject (Hull, 1970; Pociecha et al., 1988). Classification is considered as:

- the process of dividing objects in the population under study into classes, aimed at distinguishing groups of similar objects;
- the principle by which objects are included in the distinguished classes, involving the assignment of an object to a given group, in accordance with the adopted external rule;
- the division of the population under examination, the final result of which yields groups of objects most similar to one another, i.e. the result of classification activities;
- the tool and purpose of cognition.

The first attempts to employ a quantitative approach to classification issues were made by Friedrich Heincke (1898), Paul Jaccard (1900) and Jan Czekanowski (1909).

Jan Czekanowski's first diagraphic method of distance matrix ordering (1913) also gave rise to and impetus towards the development of many contemporary classification methods (Migdał-Najman & Najman, 2013). These can be broken down into various subdivisions, where the specific implementation of a given method is written in the form of an algorithm. A popular method enabling the differentiation of relatively homogeneous groups of objects is hierarchical agglomerative clustering. The outcome of the classification is presented in the form of a classification hierarchy, and graphically in the form of a dendrogram. Some of the most popular methods of agglomerative hierarchical clustering include: single-linkage clustering (Florek et al., 1951; Sneath, 1957), complete-linkage clustering (Mcquitty, 1960; Sokal & Sneath, 1963), the centroid method (Gower, 1967; Sokal & Michener, 1958), the weighted pair group method with arithmetic mean (Mcquitty, 1966, 1967), the median method (Gower, 1967; Lance & Williams, 1966), group-average linkage clustering (Sokal & Michener, 1958), and the Ward method (Ward, 1963). Hierarchical clustering is a heterogeneous procedure employing different principles (methods, criteria) for recalculating the distance metrics between objects and clusters. In 1967, G. N. Lance and W. T. Williams had propounded a schema generalising these principles, which was modified by M. Jambu in 1978 (Jambu & Lebeaux, 1978). When considering the application of hierarchical methods, the careful selection of the method used for measuring the distance between objects is of key importance. A number of such measures have been described in the literature, where the most common ones include: the Euclidean distance, the Euclidean squared distance, the Manhattan (city block / urban) distance or the Chebyshev distance. These metrics are used when the values of variables are measured on typical measurement scales. Researchers often encounter issues that are complex and consist of additive elements which jointly constitute a certain whole. Each element of this entirety can be recorded in the form of a countable or measurable quantity. If the components of this entirety are additionally expressed in the form of ratios that add up to one, then ratio-based metrics have to be used to measure the distance between the objects. One such distance measure, constituting half of the sum of the absolute proportion differences, is the measure expressed by the equation below (Balicki, 2009):

$$d_{ik} = \frac{1}{2}\sum_{j=1}^{p}\left|p_{ij} - p_{kj}\right|,$$

where $i,\ k = 1, \ldots, n,\ i \neq k$ and p_{ij} denotes the share of the j-th category (component) in the entirety of a given phenomenon characterising the i-th object, with $\sum_{j=1}^{p} p_{ij} = 1$ for each i.

This measure takes values ranging from zero to one, where zero indicates identical shares of the j-th category in the total phenomenon for the two objects compared. When a mutual decoupling of an incidence occurs, with a difference of zero for each category of the two objects compared, then the metric takes the value of one. The aim is to achieve the greatest possible separability between the distinguished classes while meeting the criteria of internal consistency of the obtained groups and ensuring their greatest possible differentiation. In empirical studies, these conditions should be checked for satisfiability. One popular procedure used to assess the grouping results is based on the MacQueen concept (1967). It presents three criteria for assessing the quality of the group structure of the analysed objects, namely the external, internal and relative criteria. In this study, the quality of the obtained group structure was assessed using the internal criterion. This approach seeks to answer the question of how well the resulting group structure, obtained by applying a given clustering technique, encapsulates the information contained in the data. The degree to which the distance matrix and the cophenetic distance matrix match can be assessed using e.g. the cophenetic correlation coefficient, presented by R. R. Sokal and F. J. Rohlf in 1962. The coefficient takes the values within the [–1; 1] range. A close match between the distance matrix and the cophenetic matrix is ensured when the coefficient takes the highest possible values, close to 1. It can be then concluded that the dendrogram, representing a concrete result of the applied hierarchical clustering strategy, provides a satisfactory enough encapsulation of the phenetic relationships. It may be further assumed that a hierarchical group structure has been created. Other indicators attributable to the internal criterion are: the Goodman-Kruskal gamma proposed in 1954 (Goodman & Kruskal, 1954), the gamma index (Baker & Hubert, 1975; Hubert, 1974), Kruskal's STRESS (Kruskal, 1964), the Dunn index (Dunn, 1974) and the popular silhouette coefficient, propounded by Rousseeuw in 1987.

Classification methods can be applied in various scientific fields and disciplines. The interdisciplinarity of data classification thus constitutes a great asset. Practical applications in this regard have been described in works employing classification methods, e.g. to filter spam from relevant emails (Migdał-Najman & Najman, 2013; Tuteja & Bogiri, 2016) or analysing Twitter posts to predict social network user personality (Pratama & Sarno, 2016).

4. Identification of Twitter user age groups

As part of the study, each of the distinguished user age group was checked for incidence of: capital letters, digits, numbers written in words, hashtags, user mentions and website links. In addition, counts of specific punctuation marks (e.g. '!' or '.'), emoticons (e.g. 'oO' or ':)') and non-relevant characters and stopwords (e.g. 'and' or 'am') were obtained. These three elements accounted for the largest portion of all

variables included in the study due to their variety/abundance: there are 10 single punctuation marks alone and used in different combinations, while emoticons created with these characters and stopwords are in the hundreds. For such prepared data, a matrix was created with 67 rows representing age groups (including 13-to 79-year-olds) and 318 columns showing the incidence of the specific features in the studied textposts. The values occurring in the matrix were then converted into relative shares reflecting the age group profiles. This matrix was used in the next step, i.e. in the cluster analysis.

The profile data in the presented study led to the application of a proportion-based distance measure (see the Formula). The cophenetic correlation coefficient was used to select the clustering method (see Table 1). The highest coefficient value was obtained for the group average method (marked in bold).

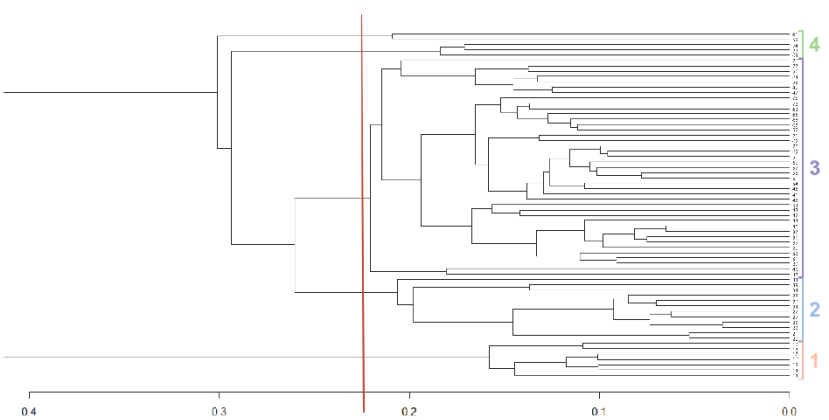
Table 1. Cophenetic coefficient values

Clustering method	Value
Nearest neighbour	0.6655
Farthest neighbour	0.8321
Group average	**0.8710**
Median	0.6795
Centroid	0.8232

Source: authors' work.

Users aged 78 and 79 formed one-element clusters, which did not merge with others, so were removed from further analysis. Based on the Mojena (1977) criterion, 4 clusters were distinguished, as presented by the dendrogram shown in Figure 4.

Figure 4. Dendrogram presenting the similarity structure of the analysed age groups



Source: authors' work.

Cluster 1 consists of people aged 13–19, i.e. the youngest Twitter users. This cluster is characterised by the highest shares of in-text numerals, capital letters and website links as well as a high share of user mentions. This indicates a greater inclination to utilise tools facilitating communication and the spread of information via Twitter. Isolated cluster 2 encompasses persons aged 20–28, 33–34 and 39 and shows the highest share of numbers written in words (e.g. 'one hundred') and the lowest share of user mentions. This cluster consists of Twitter users possibly the least concerned with publicity and retweeting messages. Cluster 3, encompassing 43 individual age groups, is the most numerous one. Its median age is 52. This cluster includes Twitter users who do not stand out significantly from the average values in terms of the use of digits, capital letters, website links, hashtags or user mentions in tweets. Finally, cluster 4 encompasses persons aged 59, 61, 69, 72 and 74, which form a relatively homogeneous group. This cluster is distinguished by the highest share of user mentions and the lowest share of website link use in tweets, which either seems to illustrate the manner in which the above-mentioned users communicate or indicates their wish to reach a larger audience.

Table 2. Basic statistics on the extracted clusters

Characteristics	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Median age	16	25.5	52	69
Average age	16	26.83	52.58	67
Share of numbers written in words in %	0.19	0.21	0.20	0.18
Share of: digits in %	0.95	0.87	0.80	0.84
capital letters in %	8.17	6.88	6.64	8.13
website links in %	0.35	0.33	0.24	0.21
hashtags in %	0.13	0.13	0.12	0.13
user mentions in %	1.28	0.92	1.17	1.46

Source: authors' work.

The basic statistics mentioned in Table 2 indicate that the use of hashtags and numerals in textposts has no significant effect on the identification of similar age groups on Twitter. The use of capital letters and in-text user mentions shows that clusters 1 and 4 are most similar to each other in this regard, as are clusters 2 and 3. The share of in-text website links use, on the other hand, indicates that clusters 1 and 2 are similar to each other, as are clusters 3 and 4.

The number of emoticons, punctuation marks and stopwords utilised by the Twitter users under examination is very extensive. Table 3 presents a ranking of the ten most commonly used items of the above-mentioned categories in each cluster. The majority of these items are recurrent for each cluster; nevertheless, certain differences in their ranking can be noted.

Table 3. Ranking of the ten most common emoticons, punctuation marks and stopwords by cluster

Ranking no.	Cluster 1	Cluster 2	Cluster 3	Cluster 4
		Emoticons		
1	:(	:)	:)	:)
2	:)	:(	d:	:(
3	:D	:3	DX	DX
4	XD	XD	:(	oO
5	:/	:/	oO	d:
6	:3	oO	XP	;)
7	:))	d:	:3	XP
8	oO	XP	XD	:3
9	d:	8D	QQ	XD
10	XP	DX	:D	8D
		Punctuation marks		
1	@	.	.	@
2	/	/	@	.
3	.	@	/	/
4	:	:	,	,
5	_	,	:	!
6	,	!	!	:
7	!	_	’	’
8	’	’	_	_
9	?	?	?	?
10	#	#	#	#
		Stopwords		
1	a	a	a	a
2	i	i	i	i
3	in	in	in	he
4	an	he	he	in
5	he	an	an	an
6	re	re	re	re
7	on	on	on	on
8	at	at	the	the
9	is	the	at	at
10	it	it	or	or

Source: authors’ work.

The group of emoticons occurring in all the clusters specified in the ranking above includes: the smiley face ‘:)’ expressing contentment, the sad face ’:(’ expressing sadness and dejection, the cat face ‘:3’ expressing childlike innocence or showing affection, the grinning face ‘d:’ expressing happiness and joy, the laughing face with closed eyes ‘XD’ expressing loud laughter, the surprised face with big eyes ‘oO’, and the playful face with a tongue sticking out and closed eyes ‘XP’ emphasizing the funny, humorous sense of a given utterance. In addition to the shared ones, emoticons individual to each cluster can be distinguished, i.e. the double chin face ‘:))’ expressing joy and a wide grin in cluster 1, and the weeping face ‘QQ’ expressing a cry caused by

embarrassment or helplessness in cluster 3. The set of emoticons specific to clusters 2 and 4 does not include emoticons which would not be shared by other clusters; nevertheless, both these groups contain emoticon '8D' expressing laughing, which is not found in the other clusters. Emoticon ':(' (sadness) is most frequently used by members of cluster 1, whose median age is 16. The 'XD' emoticon dominates in clusters 1 and 2, encompassing the young Twitter users, while emoticon 'DX' (anguish) in clusters 3 and 4.

The slash (/) punctuation mark is most prevailing in tweets posted by individuals belonging to cluster 3. The @ sign differentiates the four clusters the least. The period (.) is most commonly used by people belonging to cluster 3 and 4, while the comma (,) is used most frequently by those from cluster 4. A certain correlation characterising in-text use of punctuation marks can be observed in each individual cluster. The older the user, the more frequent the use of the comma, exclamation point and apostrophe, and the less common the use of the colon and underscore (_). This can be related to the older users' increased attention to using proper punctuation.

The ranking of stopwords shows no significant differences. In the first two clusters, the same wording can be found in the same order: 'a' and 'I'. The other stopwords occurring in all the clusters include: 'in', 'he', 'an', 're', 'on' and 'at'. Cluster 1 is distinguished by the occurrence of 'is' and the absence of 'the', as opposed to the other clusters. Stopwords differentiate the clusters to a small extent only. The most commonly used stopwords in all the clusters under examination were 'a' and 'I'; 'he' most frequently occurred in clusters 2 and 3, 'an' in cluster 1, and 'in' in cluster 4.

5. Conclusions

The elements usually eliminated from texts analysed using text mining methods do contain useful information; nevertheless, individual age groups are characterised by a relatively small variation in this regard. These elements include in particular: capital letters, numerals, numbers written in words, hashtags, user mentions or website links. The distinguished textpost elements differentiate the age groups only to a small extent. Capital letters seem to be most commonly used by the youngest (aged 13–20) and the oldest (aged 69+) Twitter users. The users aged 13–20 apply numerals more frequently than the other age groups, yet they are least likely to use numbers written in words. Persons over the age of 62 use digits less often, compared to other age groups. 39-, 42-, 53- and 63-year-olds stand out in terms of hashtag use, whereas 73-year-olds are the least likely to apply this symbol. Mentions of other users are the least frequent among the 17- to 32-year-olds, reaching the lowest level for users aged 25. Outside this group, the share of user mentions is noticeably higher. Web links are most often included in textposts of the youngest users (13- to 32-years-olds). The

share decreases for older age groups and reaches a minimum for the 59- to 62-year-olds. In the individual age groups distinguished, differences can also be noted in the structure of the emoticons used, as well as in the number of the capital letters and user mentions, relative to the number of textpost characters. The hierarchical agglomerative clustering method employed in the study allowed the identification of 4 clusters of users, which are relatively homogeneous in terms of the correctness of tweet formatting. The study indicates a general tendency in modern applied language towards the simplification of textposts. This is primarily associated with the use of pictorial language, omitting punctuation marks, the replacement of words with numbers or the use of abbreviated forms. The greatest effect of the economisation of writing can be observed among the youngest generation, which tends to use periods and commas the least frequently and is more likely to incorporate digital writing, avoiding writing numbers in words. The youngest users are also more likely to mention users and refer websites. Taking all the analysed age groups into consideration, it is the youngest who use most often simplified language in textposts.

References

Arabie, P., Hubert, L. J., & De Soete, G. (Eds). (1996). *Clustering and Classification*. World Scientific. https://doi.org/10.1142/1930.

Baker, F. B., & Hubert, L. J. (1975). Measuring the Power of Hierarchical Cluster Analysis. *Journal of the American Statistical Association*, *70*(349), 31–38. https://doi.org/10.1080/01621459.1975.10480256.

Balicki, A. (2009). *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*. Wydawnictwo Uniwerytetu Gdańskiego.

Dunn, J. C. (1974). Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics*, *4*(1), 95–104. https://doi.org/10.1080/01969727408546059.

Florek, K., Łukaszewicz, J., Perkal, J., Steinhaus, H., & Zubrzycki, S. (1951). Taksonomia wrocławska. *Przegląd Antropologiczny*, *17*, 193–211.

Goban-Klas, T. (2005). *Media i komunikowanie masowe. Teorie i analizy prasy, radia, telewizji i Internetu*. Wydawnictwo Naukowe PWN.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of Association for Cross Classifications. *Journal of the American Statistical Association*, *49*(268), 732–764. https://doi.org/10.1080/01621459.1954.10501231.

Gower, J. C. (1967). A Comparison of Some Methods of Cluster Analysis. *Biometrics*, *23*(4), 623–637. https://doi.org/10.2307/2528417.

Hubert, L. (1974). Approximate evaluation techniques for the single-link and complete-link hierarchical clustering procedures. *Journal of the American Statistical Association*, *69*(347), 698–704. https://doi.org/10.1080/01621459.1974.10480191.

Hull, D. L. (1970). Contemporary Systematic Philosophies. *Annual Review of Ecology, Evolution, and Systematics*, *1*(1), 19–54. https://doi.org/10.1146/ANNUREV.ES.01.110170.000315.

Jakobson, R. (1960). Poetyka w świetle językoznawstwa. *Pamiętnik Literacki*, *51*(2), 431–473.

Jambu, M., & Lebeaux, M. O. (1978). *Classification automatiqe pour l'analyse des donnees: vol. 1. Méthodes et algorithms*. Paris Dunod.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*(2), 115–129. https://doi.org/10.1007/BF02289694.

Lance, G. N., & Williams, W. T. (1966). A Generalized Sorting Strategy for Computer Classifications. *Nature*, *212*, 218. https://doi.org/10.1038/212218a0.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. E. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: vol. 1. Statistics* (pp. 281–298). University of California Press. https://projecteuclid.org/proceedings/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Fifth-Berkeley-Symposium-on-Mathematical-Statistics-and/Chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bsmsp/1200512992.

Majkowska, A., Migdał-Najman, K., Najman, K., & Raca, K. (2021). Identification of the Words Most Frequently Used by Different Generations of Twitter Users. In K. Jajuga, K. Najman & M. Walesiak (Eds.), *Data Analysis and Classification. Methods and Applications* (pp. 27–47). Springer. https://doi.org/10.1007/978-3-030-75190-6_3.

Majkowska, A., Migdał-Najman, K., Najman, K., & Raca, K. (2022). Graphic Characters as Twitter Age Group Identifiers. In K. Jajuga, G. Dehnel & M. Walesiak (Eds.), *Modern Classification and Data Analysis. Methodology and Applications to Micro- and Macroeconomic Problems* (pp. 275–288). Springer, Cham. https://doi.org/10.1007/978-3-031-10190-8_19.

Mcquitty, L. L. (1960). Hierarchical Linkage Analysis for the Isolation of Types. *Educational and Psychological Measurement*, *20*(1), 55–67. https://doi.org/10.1177/001316446002000106.

Mcquitty, L. L. (1966). Similarity Analysis by Reciprocal Pairs for Discrete and Continuous Data. *Educational and Psychological Measurement*, *26*(4), 825–831. https://doi.org/10.1177/001316446602600402.

Mcquitty, L. L. (1967). Expansion of Similarity Analysis By Reciprocal Pairs for Discrete and Continuous Data. *Educational and Psychological Measurement*, *27*(2), 253–255. https://doi.org/10.1177/001316446702700202.

Migdał-Najman, K., & Najman, K. (2013). *Samouczące się sztuczne sieci neuronowe w grupowaniu i klasyfikacji danych. Teoria i zastosowania w ekonomii*. Wydawnictwo Uniwersytetu Gdańskiego.

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: An evaluation. *The Computer Journal*, *20*(4), 359–363. https://doi.org/10.1093/COMJNL/20.4.359.

Pociecha, J., Podolec, B., Sokołowski, A., & Zając, K. (1988). *Metody taksonomiczne w badaniach społeczno-ekonomicznych*. Państwowe Wydawnictwo Naukowe.

Pratama, B. Y., & Sarno, R. (2016). Personality classification based on Twitter text using Naive Bayes, KNN and SVM. In *Proceedings of 2015 International Conference on Data and Software Engineering* (pp. 170–174). Universitas Gadjah Mada. https://doi.org/10.1109/ICODSE.2015.7436992.

Rodrigues, D., Prada, M., Gaspar, R., Garrido, M. V., & Lopes, D. (2018). Lisbon Emoji and Emoticon Database (LEED): Norms for emoji and emoticons in seven evaluative dimensions. *Behavior Research Methods*, *50*(1), 392–405. https://doi.org/10.3758/S13428-017-0878-6.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Sneath, P. H. A. (1957). The Application of Computers to Taxonomy. *Journal of General Microbiology*, *17*(1), 201–226. https://doi.org/10.1099/00221287-17-1-201.

Sokal, R. R., & Michener, C. D. (1958). A Statistical Method for Evaluating Systematic Relationships. *The University of Kansas Science Bulletin*, *38*(22), 1409–1438.

Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon*, *11*(2), 33–40. https://doi.org/10.2307/1217208.

Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of Numerical Taxonomy*. W. H. Freeman & Company.

Sztompka, P. (2002). *Socjologia. Analiza społeczeństwa*. Znak.

Tuteja, S. K., & Bogiri, N. (2016). Email Spam filtering using BPNN classification algorithm. In *Proceedings of 2016 International Conference on Automatic Control and Dynamic Optimization Techniques* (pp. 915–919). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICACDOT.2016.7877720.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845.

Wyka, K. (1939). Rozwój problemu pokolenia. *Przegląd Socjologiczny*, *7*(1–2), 159–192.

Wyka, K. (1977). *Pokolenia literackie*. Wydawnictwo Literackie.

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(10), 75–88 https://doi.org/10.59139/ws.2024.10.5
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

The role of Artificial Intelligence when generating official statistical data

Rola sztucznej inteligencji w generowaniu oficjalnych danych statystycznych

1. Introduction

In today's data-rich world, the field of statistics has undergone a profound transformation, adapting to the complexities and challenges posed by modern data environments. The rise of big data offers opportunities but also poses challenges for statisticians. Traditional statistical methods have proven insufficient in terms of managing the scale and complexity of the constantly increasing volumes of datasets. Modern statistics, however, has adapted to these changes through scalable algorithms, distributed computing frameworks and parallel processing techniques to analyse massive datasets efficiently (Yung et al., 2018). Moreover, statisticians emphasize the importance of data pre-processing, dimensionality reduction and feature engineering to yield meaningful insights from large and complex data sources (Chu & Poirier, 2015; United Nations Economic Commission for Europe [UNECE], 2021; Yung et al., 2018).

In the digital age, the production of official statistics is undergoing a transformative evolution thanks to the integration of Artificial Intelligence (AI) technologies. From data collection to analysis and reporting, AI has been playing a pivotal role in revolutionising how official statistics are generated.

The integration of AI in official statistics has transformed traditional approaches to data collection, processing and analysis. AI algorithms have enabled automated data collection from diverse sources such as surveys, administrative records and sensor networks, enhancing the efficiency of statistical data collection processes. Furthermore, machine learning techniques have been instrumental in automating data processing tasks, facilitating the production of valuable insights from vast and heterogeneous datasets (Chu & Poirier, 2015; UNECE, 2021; Yung et al., 2018).

Quality control and assurance guarantee the reliability and integrity of official statistics. AI-powered anomaly detection algorithms play a critical role in identifying outliers, errors and inconsistencies within datasets, enabling statisticians to rectify discrepancies and uphold data quality standards (Boukerche et al., 2020). Additionally, Alghushairy et al. (2020) highlight the importance of AI in detecting unusual patterns or discrepancies, allowing for proactive interventions to maintain the credibility of statistical outputs.

Predictive analytics has emerged as a cornerstone of modern statistical practice, enabling statisticians to forecast future trends and make informed decisions based on historical data. Survey optimisation and design represent another area where AI technologies have made significant contributions to generating official statistics. Machine learning algorithms optimise survey designs by identifying relevant variables, minimising respondent burden and improving response rates (Braaksma & Offermans, 2021). AI-driven approaches reduce the costs and improve the efficiency of surveys, ensuring the representativeness and reliability of statistical estimates.

In addition to enhancing efficiency and accuracy, AI also plays a crucial role in addressing ethical and privacy concerns in the production of official statistics. The legal and ethical implications of using AI and machine learning techniques in statistical production emphasise the importance of transparency, accountability and fairness. The application of differential privacy techniques protects individuals' privacy while enabling the analysis and dissemination of aggregated statistics.

In conclusion, the application of AI in official statistics represents a paradigm shift in collecting, processing and analysing data. By harnessing the power of AI, statistical agencies can improve the efficiency, accuracy and timeliness of statistical production processes, ultimately empowering evidence-based decision-making and policy formulation.

Processing data into actionable information is the core function of National Statistical Offices (NSOs) around the world, with their role being critical in shaping national and international policy decisions. Governed by the Fundamental Principles of Official Statistics (FPOS), set forth by the United Nations Statistical Commission, NSOs are tasked with capturing, analysing and disseminating official national statistics ensuring their reliability and international comparability (United Nations [UN], 2014). However, as technological advancements continue to redefine data collection and analysis methods, NSOs face the challenge of integrating new technologies while upholding the principles of accuracy, reliability and impartiality.

Traditionally, NSOs have relied on labour-intensive methods for generating official statistics. However, the exponential growth of data, coupled with the emergence of AI technologies, necessitates a shift towards more machine-intensive and automated strategies (Chu & Poirier, 2015; Julien et al., 2020; UNECE, 2021). Machine learning algorithms unlock the potential to streamline tasks such as data editing, imputation, categorisation and coding.

Nevertheless, while adopting AI technologies, NSOs are faced with a set of complex considerations. On the one hand, AI-driven automation of statistical tasks promises real-time insights and enhanced productivity which allows NSOs to keep pace with the ever-increasing demands for timely data. On the other hand, concerns regarding the accuracy, reliability and interpretability of AI-generated outputs must be carefully addressed. The trade-off between real-time information and the quality of statistical

outputs is a critical issue that NSOs must navigate as they integrate AI into their operations (Chu & Poirier, 2015; Julien et al., 2020; UNECE, 2021).

Furthermore, the data infrastructure, institutional capacity and regulatory frameworks which shape the technological development and strategies of NSOs vary across countries. These factors can influence the adoption and implementation of AI technologies. Understanding the differences and similarities between NSOs worldwide provides significant insights essential for developing data-driven strategies in the context of the digital transformation in the public sector.

In conclusion, NSOs play a pivotal role in processing data into actionable information of a national and international significance. When adopting machine-intensive and automation strategies based on AI technologies, NSOs must balance the pursuit for real-time information with the principles of accuracy, reliability and impartiality to harness the transformative potential of AI while upholding their mandate to deliver high-quality official statistics.

2. AI integration and the Fundamental Principles of Official Statistics

The integration of AI into the realm of official statistics entails both opportunities and challenges. As NSOs navigate this technological transformation, they must uphold the FPOS while harnessing the potential of AI.

One of the FPOS is impartiality, ensuring that statistics are produced without political or ideological influence. AI contributes to impartiality by automating processes, minimising human intervention and reducing the potential for bias. For example, AI algorithms can standardise data collection methods, eliminating subjective decision-making and ensuring consistency across data sources (United Nations Statistics Division [UNSD], n.d.). By removing human biases, AI helps maintain the objectivity of statistical analyses, reinforcing the credibility of official statistics.

Another FPOS is reliability which states that data should be accurate, consistent and free from errors. AI enhances reliability by automating quality control processes and detecting anomalies within datasets. Machine learning algorithms can identify inconsistencies or outliers that might indicate data errors, enabling statisticians to promptly address any issues (UNSD, n.d.). Through continuous learning and adaptation, AI systems improve the accuracy and reliability of statistical outputs, bolstering confidence in their integrity.

Official statistics must remain relevant and timely to meet the evolving needs of policymakers, businesses and the public. AI enables statisticians to analyse vast amounts of data in a short time, providing nearly real-time insights. Predictive analytics powered by AI can forecast future trends, enabling proactive decision-making and policy formulation (UNSD, n.d.). Moreover, AI facilitates the integration of diverse data sources, enriching official statistics with contextual information and enhancing their relevance to stakeholders.

Transparency and accountability are FPOS that underpin the credibility of official statistics. While AI introduces complexity into statistical processes, it also enables greater transparency through documentation and traceability. AI algorithms, data sources and decision-making processes can be documented, ensuring transparency and reproducibility (UNSD, n.d.). Additionally, accountability mechanisms can be implemented to monitor AI systems and address any biases or errors that may arise during statistical production.

Ethical considerations are paramount in the integration of AI into official statistics, particularly concerning privacy protection and data confidentiality. AI techniques such as differential privacy can be employed to anonymise data while preserving its utility for statistical analysis. Moreover, statisticians must adhere to ethical guidelines and regulations to ensure the responsible use of AI in statistical production, safeguarding individuals' rights and maintaining public trust.

As AI becomes increasingly integrated into official statistics, it is essential to uphold the FPOS of impartiality, reliability, relevance, transparency and accountability. By leveraging AI technologies responsibly, statisticians can enhance the quality, integrity and timeliness of official statistics while addressing emerging challenges in data collection, analysis and dissemination. Through collaboration, innovation and adherence to ethical standards, AI and official statistics can work synergistically to provide valuable insights and support evidence-based decision-making in a data-driven world.

3. AI in official statistics

The role of AI in official statistics is multifaceted and continuously evolving. Below are some key aspects of AI's contribution to official statistics.

3.1. Data collection and processing

AI technologies such as natural language processing (NLP), computer vision and machine learning algorithms are used to automate processes of data collection from various sources, including administrative records, surveys, sensor networks and social media. These technologies can prove helpful in extracting relevant information from unstructured data sources and in streamlining data processing tasks (Barrachina et al., 2009; Koch, 2016; Silver et al., 2018). Official statistics uses AI methods for classification as well as for recognition, estimation and/or the imputation of relevant characteristic values of statistical units (UNECE, 2021).

One of the key advantages of AI in data collection is its ability to automate the extraction of information from unstructured data sources, such as text documents, images and videos. NLP techniques enable AI algorithms to parse and analyse textual

data, extract key insights and categorise information based on predefined criteria. For example, AI-powered chatbots can engage with respondents in natural language conversations to collect survey responses or gather information from social media posts and online forums. Similarly, computer vision algorithms can analyse images and videos to identify relevant data points, such as demographic information or spatial features, enabling statistical agencies to incorporate multimedia data into their analyses. Moreover, machine learning algorithms can be trained on historical data to classify observations into predefined categories, such as industry sectors, occupation codes or demographic groups. This automated classification process not only accelerates data processing but also reduces the potential for human error and bias. Additionally, AI algorithms can perform recognition tasks, such as optical character recognition (OCR) for digitising paper-based documents or speech recognition for transcribing audio recordings, further streamlining data collection and processing workflows. AI methods also enable statistical agencies to leverage alternative data sources and innovative data collection techniques, such as web scraping, sentiment analysis and social media mining. By harnessing the vast amounts of digital data generated by online platforms and digital devices, AI-powered data collection methods complement traditional survey-based approaches, providing additional insights into societal trends, consumer behaviour and economic indicators. For example, sentiment analysis algorithms can analyse social media posts to gauge public sentiment towards specific topics or products, providing real-time indicators of public opinion and behaviour.

3.2. Quality control and error detection

As datasets continue to grow in size and complexity, the task of identifying and correcting data errors becomes increasingly challenging for NSOs and other statistical agencies. Traditional manual approaches to quality control are often time-consuming and labour-intensive, involving a thorough inspection of every data point. In this context, AI techniques offer valuable tools for automating quality control processes and detecting anomalies, such as data errors, outliers and inconsistencies in large and diverse datasets. AI algorithms can identify patterns and deviations through such techniques as clustering, classification and regression. As a result, statisticians are able to prioritise investigation and corrective action, ultimately enhancing the quality and reliability of statistical outputs (Alghushairy et al., 2020; Boukerche et al., 2020).

Furthermore, AI techniques enable predictive modelling approaches that anticipate and pre-emptively address data quality issues before they propagate through the analytical process. By training models on historical data and learning from past errors, AI algorithms can predict the likelihood of future data anomalies and proactively

implement preventive measures. For example, predictive models can detect abnormal data patterns, missing values or inconsistencies and automatically trigger alerts or corrective actions, such as data imputation or outlier removal.

3.3. Forecasting and predictive analytics

Forecasting and predictive analytics provide insights into future trends and potential outcomes, allowing for effective policy decision-making and strategic planning. In the realm of official statistics, machine learning models used for forecasting have emerged as powerful tools applied for predicting accurate and reliable values of socio-economic indicators. By analysing historical data patterns, relationships and trends, machine learning models, as opposed to traditional statistical methods, can generate forecasts for key indicators such as GDP growth, employment rates, inflation, population trends and more (Aburto & Weber, 2007; Wickramasuriya et al., 2019). Machine learning offers a diverse array of modelling techniques that can be tailored to different forecasting tasks and data types. From traditional time series analysis methods like ARIMA (Auto Regressive Integrated Moving Average) to more advanced algorithms such as neural networks, ensemble methods and gradient boosting, machine learning provides statisticians with a versatile toolkit for building forecasting models that can manage complex data structures and non-linear relationships (Aburto & Weber, 2007).

One of the key advantages of machine learning-based forecasting is its ability to provide real-time or near-real-time predictions. By continuously updating models with the latest data, machine learning algorithms can generate forecasts that reflect the most current trends and developments, enabling policymakers to make timely and informed decisions. Real-time forecasting is particularly valuable in fast-paced environments where rapid responses to changing conditions are essential (Wickramasuriya et al., 2019). Moreover, machine learning-based forecasting enables planning and risk management analysis by simulating various future scenarios and assessing their potential impact on socio-economic indicators. These include forecasting economic growth, predicting unemployment rates, or projecting demographic trends (Wickramasuriya et al., 2019). Machine learning models thus help policymakers evaluate different policy options, anticipate potential risks and develop contingency plans to mitigate adverse outcomes (Aburto & Weber, 2007).

3.4. Survey optimisation and design

The integration of AI algorithms into survey design represents a transformative leap in the field of official statistics, offering unparalleled opportunities for the optimisation of data collection processes. With the ever-growing complexity of data and the need for accurate timely insights, AI-driven approaches are revolutionising

the way surveys are designed, conducted and analysed. This integration not only enhances the efficiency and effectiveness of survey methodologies but also improves the quality and reliability of statistical outputs.

One of the key advantages of AI-driven survey optimisation is its ability to streamline the selection of variables and determination of sample sizes. Traditional survey design processes often rely on manual selection methods, which can be time-consuming and prone to bias. AI algorithms, on the other hand, can analyse large datasets to identify the relevant variables and predictors that influence survey outcomes. By leveraging techniques such as feature selection and predictive modelling, AI algorithms can identify the most informative variables to include in the survey instruments, ensuring that the surveys capture the essential information while minimising respondent burden. Additionally, AI algorithms can determine optimal sample sizes and sampling techniques based on the desired levels of precision and statistical power, allowing statisticians to design surveys that achieve reliable and representative results (Cateni et al., 2013).

Another significant benefit of AI-driven survey design is the ability to create personalised survey instruments adjusted to individual respondent characteristics. Traditional surveys often follow a one-size-fits-all approach, which may not fully capture the diverse perspectives and experiences of respondents. AI-driven survey tools, however, analyse demographic data, past survey responses and other relevant information to tailor the survey questions, formats and delivery methods to individual respondents. This personalised approach increases engagement and response rates.

Furthermore, AI algorithms enable an adaptive survey design, allowing survey instruments to evolve in real-time based on the incoming responses and feedback. Typically, traditional surveys have a static structure, which may not account for changes in respondent behaviour or any emerging trends. AI-driven survey tools, on the other hand, can dynamically adjust the survey content, skip patterns and question sequencing to optimise the respondent experience and data quality. Adaptive survey design minimises respondent fatigue, reduces survey dropout rates and maximises the efficiency of data collection efforts.

In addition to enhancing survey design and data collection processes, AI-driven approaches also strengthen quality control measures and response validation mechanisms. NLP techniques enable AI algorithms to analyse open-ended responses, identify sentiment and detect semantic inconsistencies. By automating response validation processes, AI-driven survey tools streamline data cleaning efforts and improve the overall quality of the survey results.

Moreover, AI-driven predictive analytics play a crucial role in non-response adjustment. Machine learning algorithms analyse demographic, behavioural and contextual data to predict response propensity and identify factors associated with non-response bias. Predictive analytics enable statisticians to adjust the survey weights and account for non-response bias, so that the survey results accurately reflect the characteristics of the target population.

3.5. Privacy preservation and data confidentiality

The preservation of privacy and data confidentiality are one of the main concerns in official statistics. The need to protect individual-level data is balanced with the imperative to derive meaningful insights from aggregate data. This aim is achieved through differential privacy and other privacy-preserving techniques. By leveraging AI-based methods for privacy protection, statisticians can effectively address these concerns while still harnessing the power of data.

Differential privacy has emerged as the leading approach for privacy preservation in official statistics. This framework provides a rigorous mathematical definition of privacy, ensuring that the inclusion or exclusion of any single individual's data does not significantly impact the outcome of statistical analyses. By adding carefully calibrated noise to query responses or by perturbing data during collection, the differential privacy technique allows statisticians to share and analyse sensitive information without compromising individual privacy. AI-based methods further enhance privacy preservation efforts by offering sophisticated techniques for anonymisation, encryption and data obfuscation. For example, advanced encryption algorithms can secure data both at rest and in transit, preventing unauthorised access and ensuring data confidentiality throughout its lifecycle. Similarly, anonymisation techniques such as k-anonymity and l-diversity help protect against re-identification attacks by aggregating or generalising individual attributes while not distorting the analysis at the aggregate level.

Furthermore, AI-powered data masking techniques enable statisticians to generate synthetic data that closely mimic the statistical properties of the original dataset while preserving privacy. By training generative models on the original data distribution, AI algorithms can generate synthetic data samples that maintain the key statistical characteristics without revealing sensitive information about the individual respondents. This allows for a robust analysis and modelling while mitigating privacy-related risks associated with sharing or disclosing sensitive data.

In addition to these technical approaches, privacy preservation in official statistics also relies on robust governance frameworks and regulatory safeguards. NSOs and other statistical agencies adhere to strict data protection regulations and ethical guidelines to ensure compliance with privacy laws and standards. This includes implementing rigorous data access controls, conducting privacy impact assessments and establishing protocols for data sharing and dissemination that prioritise privacy and confidentiality (El Mestari et al., 2024).

3.6. Automation of reporting and dissemination

Automation of reporting and dissemination processes is a transformative practice in official statistics which has revolutionised the collection, processing, analysis and dissemination of data. Advanced technologies, including AI, ensure that statistical

agencies can streamline workflows, enhance the efficiency and accuracy of statistical information, and guarantee its timely delivery and easy access.

One of the primary benefits of automation in reporting and dissemination is that the production and dissemination of statistical outputs is fast. Traditionally, the process of collecting, processing and analysing data for reporting purposes tends to be time-consuming and labour-intensive. However, with automation technologies, including AI-powered data processing and analysis tools, statistical agencies can significantly reduce the time and effort required to generate reports and disseminate information. Automated algorithms can ingest raw data, perform complex analyses and prepare reports and visualisations in a fraction of the time it would take using manual methods. Moreover, automation allows for greater standardisation and consistency in reporting practices. By implementing predefined templates, automated reporting systems ensure that statistical outputs adhere to the established formatting guidelines and quality standards, reducing the risk of errors and inconsistencies in the reporting process, adding to the credibility and reliability of official statistics.

Furthermore, automation facilitates the customisation and personalisation of statistical information to meet the diverse needs of users. To enhance user engagement and satisfaction, AI-driven algorithms analyse user preferences, behaviour and feedback and tailor reporting and dissemination strategies accordingly. For example, statistical agencies can use machine learning algorithms to segment users based on their interests, expertise and information needs, delivering customised reports, dashboards and interactive data visualisations that are relevant and actionable. Additionally, automation enables statistical agencies to expand the accessibility and reach of official statistics through digital platforms and online portals. By using AI-driven technologies, such as natural language processing and chatbots, agencies can automate the generation of data summaries, FAQs and explanatory notes, making statistical information more understandable and accessible to non-expert users. Moreover, automated dissemination systems can deliver statistical updates and alerts in real-time via email, social media and mobile applications.

3.7. Ethical and legal considerations

As AI is becoming more pervasive in official statistics, it is essential to address ethical and legal considerations surrounding data privacy, algorithmic bias, transparency and accountability. Statisticians must ensure that AI-driven approaches comply with relevant regulations and guidelines while upholding principles of fairness, transparency and user trust.

Although AI holds a significant promise for enhancing the efficiency, accuracy and accessibility of official statistics, it also poses challenges that need to be carefully

addressed to realise its full potential in this domain. The use of AI in processing large datasets, especially those containing personal information, raises significant ethical concerns. Issues such as privacy, consent and the potential for surveillance need to be addressed. Without stringent ethical guidelines, the use of AI in official statistics could lead to privacy violations or the misuse of personal data. This could result in a loss of public trust and legal challenges.

For this purpose, policymakers and legal experts need to collaborate with statisticians and data scientists to develop regulations and ethical guidelines governing the use of AI in official statistics. This includes addressing issues like data privacy, consent, algorithmic transparency and accountability.

3.8. Geospatial analysis and spatial modelling

Geospatial analysis and spatial modelling, powered by AI techniques such as geographic information systems (GIS), spatial analysis and remote sensing are indispensable tools in the realm of official statistics. These methods play a vital role in analysing spatial patterns, conducting regional assessments and modelling spatial relationships, thereby providing crucial insights into geographical disparities, urbanisation trends, environmental impacts and other spatially explicit phenomena. By integrating AI-driven geospatial techniques into statistical analysis, official statistical agencies can support spatial planning and policy-making with comprehensive, data-driven assessments.

One of the primary applications of AI-driven geospatial analysis in official statistics is the identification and analysis of spatial patterns and trends. By leveraging GIS and spatial analysis techniques, statistical agencies can analyse spatially distributed data, such as population demographics, land use and socio-economic indicators to identify clusters, hotspots and trends across geographic regions. This enables policymakers to gain a deeper understanding of spatial disparities and inequalities, guiding targeted interventions and resources allocation strategies to address socio-economic challenges and promote inclusive development.

Moreover, AI-driven geospatial analysis facilitates the integration of diverse datasets from multiple sources, including satellite imagery, aerial photography and ground-based sensors. By combining spatial data with statistical information such as census data and survey results, statistical agencies can conduct comprehensive spatial analyses to assess the impact of urbanisation, environmental changes and natural disasters on local communities and ecosystems. This interdisciplinary approach enables policymakers to develop evidence-based policies and interventions that address complex spatial challenges such as climate change adaptation, disaster risk management and sustainable urban development.

Furthermore, AI-powered spatial modelling techniques enable statistical agencies to simulate and forecast spatial relationships and phenomena, providing valuable

insights into future scenarios and potential outcomes. By using machine learning algorithms and predictive modelling approaches, statistical agencies can analyse historical spatial data to forecast future trends such as population growth, land use changes and infrastructure development. This enables policymakers to anticipate spatial dynamics and plan for future challenges such as urban sprawl, transportation congestion and environmental degradation, ensuring sustainable and resilient spatial development.

In addition to informing policy-making at the national and regional levels, AI-driven geospatial analysis also supports local decision-making and community development initiatives. By providing spatially disaggregated data and interactive mapping tools, statistical agencies empower local authorities, non-governmental organisations and community stakeholders to identify local priorities, plan targeted interventions and monitor progress towards sustainable development goals (SDGs). This bottom-up approach to spatial planning and decision-making ensures that policies and interventions are context-specific, participatory and responsive to the needs of local communities (Hu et al., 2019; Janowicz et al., 2020; Openshaw, 1992).

4. Conclusions

The integration of AI into official statistics represents a transformative shift in how data is collected, processed, analysed and disseminated. This paradigm shift is driven by the recognition of AI's ability to tackle the complexities and challenges posed by modern data environments, including the exponential growth of data volumes and the need for real-time insights.

Through AI-driven approaches, statistical agencies can improve the efficiency, accuracy and timeliness of statistical production processes. AI technologies enable automated data collection from diverse sources, automated data processing tasks and enhanced quality control mechanisms. Additionally, AI facilitates predictive analytics, survey optimisation, and design, thereby empowering evidence-based decision-making and policy formulation.

However, the adoption of AI in official statistics also presents complex considerations, including ethical, legal and privacy concerns. As AI is becoming more pervasive, it is crucial for statistical agencies to uphold principles of impartiality, reliability, relevance, transparency and accountability. By leveraging AI responsibly and ethically, statistical agencies can maintain public trust and confidence in official statistics while harnessing the full potential of AI to address emerging challenges and opportunities in the digital age.

In addition to the FPOS and ethical considerations discussed, the integration of AI in official statistics offers several notable advantages. Firstly, AI-driven approaches enable statisticians to process and analyse vast amounts of data more efficiently and

accurately than through the application of traditional methods. This capability is particularly beneficial in the era of big data where the volume, velocity and variety of data sources continue to expand exponentially. By leveraging machine learning algorithms and other AI techniques, statisticians may uncover valuable insights, identify patterns and make predictions that were previously inaccessible or impractical.

Moreover, through AI technologies statistical agencies are able to adapt quickly to the evolving data sources and analytical methodologies. With the proliferation of new data types such as social media data, sensor data and satellite imagery, traditional statistical methods may no longer suffice to capture the complexity and nuances of modern datasets. AI-driven approaches offer flexible and adaptable solutions that can accommodate diverse data sources and analytical requirements.

Furthermore, AI fosters innovation and creativity in official statistics by empowering statisticians to explore new methods, techniques and applications. The interdisciplinary nature of AI, which draws from fields such as computer science, mathematics and cognitive psychology, encourages collaboration and cross-pollination of ideas. This interdisciplinary approach spurs innovation in statistical modelling, data visualisation and decision support systems, leading to the development of novel tools and methodologies.

Additionally, AI-driven approaches improve accessibility and usability of statistical information by making data more understandable, interactive and user-friendly. Through data visualisation techniques, NLP interfaces and interactive dashboards, statistical agencies can present complex information in intuitive and engaging formats that cater to diverse audiences. This democratisation of data empowers policymakers, researchers, businesses and the general public to access, understand and utilise official statistics effectively.

Overall, the integration of AI in official statistics represents a transformative shift that holds an immense potential to advance the field and address the evolving needs of stakeholders in an increasingly data-driven world. By embracing AI responsibly, statisticians can leverage its capabilities to produce high-quality, timely and actionable insights that drive evidence-based decision-making, promote transparency and accountability, and ultimately contribute to the development of the society as a whole.

References

Aburto, L., & Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing*, *7*(1), 136–144. https://doi.org/10.1016/j.asoc.2005.06.001.

Alghushairy, O., Alsini, R., Ma, X., & Soule, T. (2020). A Genetic-Based Incremental Local Outlier Factor Algorithm for Efficient Data Stream Processing. In *ICCDA 2020. Proceedings of the 4th International Conference on Computational Data and Analytics* (pp. 38–45). Association for Computing Machinery. https://doi.org/10.1145/3388142.3388160.

Barrachina, S., Bender, O., Casacuberta, F., Civera, J., Cubel, E., Khadivi, S., Lagarda, A., Ney, H., Tomás, J., Vidal, E., & Vilar, J. M. (2009). Statistical Approaches to Computer-Assisted Translation. *Computational Linguistics*, *35*(1), 3–28. https://doi.org/10.1162/coli.2008.07-055-r2-06-29.

Boukerche, A., Zheng, L., & Alfandi, O. (2020). Outlier detection: Methods, models, and classification. *ACM Computing Surveys*, *53*(3), 1–37. https://doi.org/10.1145/3381028.

Braaksma, B., & Offermans, M. (2021). Statistics Netherlands and AI. In *Proceedings of Statistics Canada Symposium: Adopting Data Science in Official Statistics to Meet Society's Emerging Needs*. https://www150.statcan.gc.ca/n1/en/pub/11-522-x/2021001/article/00011-eng.pdf?st=ncDUTkRG.

Cateni, S., Vannucci, M., Vannocci, M., & Colla, V. (2013). Variable Selection and Feature Extraction Through Artificial Intelligence Techniques. In L. Freitas, & A. P. B. Rodrigues De Freitas (Eds.), *Multivariate Analysis in Management, Engineering and the Sciences* (pp. 103–118). https://doi.org/10.5772/53862.

Chu, K., & Poirier, C. (2015). *Machine Learning Documentation Initiative*. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2015/Topic3_Canada_paper.pdf.

El Mestari, S. Z., Lenzini, G., & Demirci, H. (2024). Preserving data privacy in machine learning systems. *Computers & Security*, *137*, 1–22. https://doi.org/10.1016/j.cose.2023.103605.

Hu, Y., Li, W., Wright, D., Orhun, A., Wilson, D., Maher, O., & Raad, M. (2019). Artificial Intelligence Approaches. In J. P. Wilson (Ed.), *The Geographic Information Science and Technology Body of Knowledge* (pp. 1–12). University Consortium for Geographic Information Science. https://doi.org/10.22224/gistbok/2019.3.4.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, *34*(4), 625–636. https://doi.org/10.1080/13658816.2019.1684500.

Julien, C., Choi, I., Deeben, E., Yung, W., Wesley, Y., & Measure, A. (2020). *HLG-MOS Machine Learning Project*. https://statswiki.unece.org/display/ML/HLG-MOS+Machine+Learning+Project.

Koch, C. (2016). How the Computer Beat the Go Player. *Scientific American Mind*, *27*(4), 20–23. https://doi.org/10.1038/scientificamericanmind0716-20.

Openshaw, S. (1992). Some Suggestions Concerning the Development of Artificial Intelligence Tools for Spatial Modelling and Analysis in GIS. *The Annals of Regional Science*, *26*(1), 35–51. https://doi.org/10.1007/BF01581479.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*(6419), 1140–1144. https://doi.org/10.1126/science.aar6404.

United Nations. (2014). Resolution adopted by the General Assembly on 29 January 2014. Fundamental Principles of Official Statistics (A/RES/68/261). https://unstats.un.org/unsd/dnss/gp/fp-new-e.pdf.

United Nations Economic Commission for Europe. (2021). *Machine Learning for Official Statistics*. United Nations. https://unece.org/sites/default/files/2022-09/ECECESSTAT20216.pdf.

United Nations Statistics Division. (n.d.). *Principles Governing International Statistical Activities*. https://unstats.un.org/unsd/ccsa/principles_stat_activities/.

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of American Statistical Association*, *114*(526), 804–819. https://doi.org/10.1080/01621459.2018.1448825.

Yung, W., Karkimaa, J., Scannapieco, M., Barcarolli, G., Zardetto, D., Ruiz Sanchez, J. A., Braaksma, B., Buelens, B., & Burger, J. (2018). *The use of machine learning in official statistics*. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2024-07/HLGMOS%20The%20use%20of%20machine%20learning%20in%20official%20statistics.pdf.

Arzu Taghiyeva
State Statistical Committee of the Republic of Azerbaijan; PhD student on Econometrics at Scientific-Research Institute of Economic Studies of Azerbaijan State University of Economics
ORCID: https://orcid.org/0000-0002-8031-4768. E-mail: arzuvalehqizi@gmail.com

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(10), 89–90
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

WYDAWNICTWA GUS. WRZESIEŃ 2024
PUBLICATIONS OF STATISTICS POLAND. SEPTEMBER 2024

W ofercie wydawniczej Głównego Urzędu Statystycznego z ubiegłego miesiąca warto zwrócić uwagę na następującą publikację:
Among Statistics Poland's publications from the previous month, we would like to recommend:

Sytuacja makroekonomiczna w Polsce na tle procesów w gospodarce światowej w 2023 r.
Macroeconomic situation in Poland in the context of the world economic processes in 2023

Dwunasta edycja wydawanej corocznie publikacji poświęconej zjawiskom makroekonomicznym i wybranym zagadnieniom społecznym. Sytuację społeczno-gospodarczą w Polsce ukazano w kontekście międzynarodowym, przede wszystkim europejskim.

Język: polski (przedmowa, spis treści, synteza, tablice i wykresy również w języku angielskim)
Language: Polish (preface, contents, summary, tables and charts available also in English)
Seria: Analizy statystyczne
Series: Statistical analyses
Dostępne wersje: drukowana i elektroniczna
Available in: printed and electronic form

Przedstawiona w opracowaniu analiza dotyczy m.in. przebiegu procesów makroekonomicznych rynku pracy, finansów publicznych, sytuacji na światowych rynkach finansowych oraz wpływu pandemii COVID-19 i wojny w Ukrainie na gospodarkę. Skoncentrowano się na opisie sytuacji w 2023 r., ale uwzględniono także procesy zachodzące w dłuższym czasie.

W publikacji wykorzystano dane z różnych źródeł. Podstawowym były badania statystyczne prowadzone przez Główny Urząd Statystyczny. Czerpano także z informacji pochodzących m.in. z Komisji Europejskiej (w tym Eurostatu), Międzynarodowego Funduszu Walutowego, Organizacji Współpracy Gospodarczej i Rozwoju, Międzynarodowej Organizacji Pracy, Konferencji Narodów Zjednoczonych ds. Handlu i Rozwoju czy Banku Światowego. Dane o sektorze finansów publicznych w Polsce uzyskano przede wszystkim z Ministerstwa Finansów.

Należy zaznaczyć, że omawiana publikacja ma charakter eksperymentalny, a zawarte w niej dane nie stanowią oficjalnych danych GUS.

We wrześniu br. ukazały się ponadto:

- *Bezrobocie rejestrowane 2 kwartał 2024 r.*;
- „Biuletyn statystyczny" nr 8/2024;
- *Budżety gospodarstw domowych w 2023 r.*;
- *Ceny robót budowlano-montażowych i obiektów budowlanych (lipiec 2024 r.)*;
- *Fizyczne rozmiary produkcji zwierzęcej w 2023 r.*;
- *Koniunktura w przetwórstwie przemysłowym, budownictwie, handlu i usługach 2000–2024 (wrzesień 2024)*;
- *Kultura i dziedzictwo narodowe w 2023 r.*;
- *Nakłady i wyniki przemysłu – 1–2 kwartał 2024 r.*;
- *Pomoc społeczna i opieka nad dzieckiem i rodziną w 2023 r.*;
- *Produkcja ważniejszych wyrobów przemysłowych w sierpniu 2024 r.*;
- *Równoległy oraz wyprzedzający zagregowany wskaźnik koniunktury, zegar koniunktury – szereg do lipca 2024 r.*;
- „Statistics in Transition new series" nr 3/2024;
- *Sytuacja demograficzna Polski do 2023 r.*;
- *Sytuacja społeczno-gospodarcza kraju w sierpniu 2024 r.*;
- *Sytuacja społeczno-gospodarcza województw Nr 2/2024*;
- *Telekomunikacja 2023*;
- *Transport – wyniki działalności w 2023 r.*;
- „Wiadomości Statystyczne. The Polish Statistician" nr 9/2024;
- *Zatrudnienie i wynagrodzenia w gospodarce narodowej w pierwszym półroczu 2024 r.*

Joanna Sadowy
Główny Urząd Statystyczny, Departament Opracowań Statystycznych, Polska
Statistics Poland, Statistical Products Department, Poland

DLA AUTORÓW
FOR THE AUTHORS

(for the English translation of the information given below, please visit ws.stat.gov.pl/ForAuthors)

W „Wiadomościach Statystycznych. The Polish Statistician" („WS") zamieszczane są artykuły o charakterze naukowym poświęcone teorii i praktyce statystycznej, które prezentują wyniki oryginalnych badań teoretycznych lub analitycznych wykorzystujących metody statystyki matematycznej, opisowej bądź ekonometrii. Ukazują się również artykuły przeglądowe, recenzje publikacji naukowych oraz inne opracowania informacyjne. W czasopiśmie publikowane są prace w języku polskim i angielskim.

Od 2007 r. „WS" znajdują się na liście czasopism naukowych Ministerstwa Nauki i Szkolnictwa Wyższego. Zgodnie z komunikatem Ministra Nauki z dnia 5 stycznia 2024 r. w sprawie wykazu czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych „WS" otrzymały 70 punktów.

„Wiadomości Statystyczne. The Polish Statistician" są udostępniane w następujących bazach, repozytoriach, katalogach i wyszukiwarkach: Agro, BazEkon, Biblioteka Nauki, Central and Eastern European Academic Source (CEEAS), Central and Eastern European Online Library (CEEOL), Central European Journal of Social Sciences and Humanities (CEJSH), Directory of Open Access Journals (DOAJ), EBSCO Discovery Service, European Reference Index for the Humanities and Social Sciences (ERIH Plus), Exlibris Primo, Google Scholar, ICI Journals Master List, ICI World of Journals, Norwegian Register for Scientific Journals and Publishers (The Nordic List), Summon i WorldCat.

Za publikację artykułów na łamach „WS" autorzy nie otrzymują honorariów ani nie wnoszą opłat.

1. Zgłaszanie artykułów

Prace przeznaczone do opublikowania w „WS" należy przesyłać za pośrednictwem platformy Editorial System: www.editorialsystem.com/ws.

Zgłaszany artykuł powinien być zanonimizowany, tj. pozbawiony informacji o autorze/autorach (również we właściwościach pliku), podziękowań i informacji o źródłach finansowania, a także innych informacji wskazujących na afiliację lub umożliwiających zidentyfikowanie autora. Jeżeli w pracy występują tablice, wykresy lub mapy, powinny być umieszczone w treści artykułu. Materiały graficzne, razem z danymi do nich, należy ponadto załączyć jako osobny plik / osobne pliki, najlepiej w formacie Excel. **Prosimy o niestosowanie stylów i ograniczenie formatowania do wymogów redakcyjnych.** Więcej informacji w pkt 4 *Wymogi redakcyjne*.

Razem z artykułem należy przesłać skan/zdjęcie oświadczenia o oryginalności pracy i niezłożeniu jej w innym wydawnictwie. **Załączenie oświadczenia jest warunkiem poddania pracy ocenie wstępnej i skierowania do recenzji.**

Zgłoszenie artykułu do opublikowania w „WS" oznacza zgodę na jego udostępnienie na licencji Creative Commons Uznanie autorstwa – Na tych samych warunkach 4.0 (CC BY-SA 4.0).

Autorzy mają prawo do samodzielnego umieszczania w wybranych przez siebie repozytoriach artykułu w wersji zarówno zgłoszonej do „WS", jak i zaakceptowanej do opublikowania oraz opublikowanej, z zastrzeżeniem wymogu niezwłocznego podania w repozytorium informacji o numerze „WS", w którym praca się ukazała, wraz z linkiem do niej (DOI).

2. Przebieg prac redakcyjnych

Zgłoszony artykuł jest oceniany i opracowywany w czteroetapowym procesie:

1. **Ocena wstępna**, dokonywana przez redakcję. Polega na weryfikacji: naukowego charakteru artykułu, zgodności jego tematyki z profilem czasopisma, struktury i zawartości pracy pod kątem wymogów redakcyjnych oraz oryginalności (wykrywanie programem antyplagiatowym treści zapożyczonych, a także wygenerowanych za pomocą narzędzi sztucznej inteligencji). Na jej podstawie formułowane są uwagi i zalecenia dla autora. Poprawiona/uzupełniona przez autora praca jest kierowana do recenzji. W przypadku negatywnej weryfikacji artykuł zostaje odrzucony, a autor otrzymuje decyzję wraz z uzasadnieniem.
2. **Ocena recenzentów**, dokonywana przez specjalistów w danej dziedzinie. Artykuł oceniają dwaj recenzenci spoza jednostki naukowej, przy której afiliowany jest autor; w przypadku pracy w języku angielskim co najmniej jeden recenzent jest afiliowany przy jednostce zagranicznej. W razie sprzecznych opinii dwóch recenzentów powoływany jest trzeci recenzent. Recenzenci kierują się kryteriami oryginalności i jakości opracowania zarówno w odniesieniu do treści, jak i formy artykułu.

   Autor pracy, która otrzymała dwie pozytywne oceny, wprowadza poprawki zalecane przez recenzentów i przesyła do redakcji zmodyfikowaną wersję tekstu. Jeśli pojawi się różnica zdań dotycząca zasadności proponowanych zmian, autor jest zobligowany do uzasadnienia swojego stanowiska.
3. **Ocena Kolegium Redakcyjnego** (KR), decydująca o przyjęciu pracy do publikacji. Jest dokonywana na podstawie recenzji, z uwzględnieniem opinii redaktorów tematycznego i merytorycznego. Polega m.in. na weryfikacji dokonania przez autora zmian w artykule stosownie do uwag recenzentów. KR ocenia artykuł pod względem poprawności i spójności merytorycznej oraz zaleca autorowi wprowadzenie poprawek, jeśli są one konieczne, aby praca spełniała wymogi czasopisma. Autorowi przysługuje prawo do odwołania od decyzji o niepublikowaniu artykułu. W takim przypadku powinien on skontaktować się z redakcją „WS" i przedstawić uzasadnienie. Ostateczna decyzja w tej sprawie należy do redaktora naczelnego.

   W „WS" publikowane są wyłącznie te artykuły, które otrzymają pozytywną ocenę na każdym z wymienionych etapów i zostaną poprawione przez autora zgodnie z otrzymanymi uwagami (chyba że autor przedstawi argumenty uzasadniające nieuwzględnienie danej uwagi).

   Artykuły przyjęte przez KR do publikacji są zamieszczane na stronie internetowej czasopisma w zakładce Early View, gdzie znajdują się do czasu opublikowania w konkretnym wydaniu „WS".
4. **Opracowanie redakcyjne, autoryzacja i korekta.** Artykuł zakwalifikowany do druku jest poddawany opracowaniu redakcyjnemu, a następnie – po autoryzacji – przekazywany do składu, łamania i opracowania graficznego. Następnie wykonywane są co najmniej dwie korekty wydawnicze. Autor wykonuje korektę autorską na etapie drugiej korekty wydawniczej.

   Redakcja zastrzega sobie prawo do zmiany tytułu i śródtytułów, modyfikowania tablic, wykresów i innych elementów graficznych oraz przeredagowywania treści bez naruszenia zasadniczej myśli autora.

   W przypadku odkrycia błędów w opublikowanym artykule zamieszcza się na łamach „WS" sprostowanie, a artykuł w wersji elektronicznej jest poprawiany i umieszczany na stronie internetowej „WS" ze stosownym wyjaśnieniem.

3. Zasady etyki publikacyjnej

Wszyscy uczestnicy procesu wydawniczego: autorzy, Rada Naukowa, Kolegium Redakcyjne, pracownicy Wydziału Czasopism Naukowych w Departamencie Opracowań Statystycznych Głównego Urzędu Statystycznego, recenzenci i wydawca są zobowiązani do przestrzegania zasad etyki publikacyjnej. Zasady obowiązujące w „Wiadomościach Statystycznych. The Polish Statistician" („WS") opierają się na wytycznych Komitetu ds. Etyki Publikacyjnej (Committee on Publication Ethics – COPE), które są dostępne na stronie internetowej http://www.publicationethics.org.

3.1. Odpowiedzialność autorów

1. Artykuły naukowe nadsyłane do „WS" powinny zawierać precyzyjny opis badanych zjawisk i stosowanych metod oraz autorskie wnioski i sugestie dotyczące rozwoju badań i analiz statystycznych. Autorzy ponoszą odpowiedzialność za treści prezentowane w artykułach oraz właściwe cytowanie prac innych autorów. W razie zgłaszania przez czytelników zastrzeżeń odnoszących się do tych treści autorzy są zobligowani do udzielenia odpowiedzi za pośrednictwem redakcji.
2. Na autorach spoczywa obowiązek zapewnienia pełnej oryginalności przedłożonych prac. Redakcja nie toleruje przejawów nierzetelności naukowej autorów, takich jak:
   - duplikowanie publikacji – ponowne publikowanie własnego utworu lub jego części;
   - plagiat – przywłaszczenie cudzego utworu lub jego fragmentu bez podania informacji o źródle;
   - fabrykowanie danych – oparcie pracy naukowej na nieprawdziwych wynikach badań;
   - autorstwo widmo (*ghost authorship*) – nieujawnianie współautorów, mimo że wnieśli oni istotny wkład w powstanie artykułu;
   - autorstwo gościnne (*guest authorship*) – podawanie jako współautorów osób o znikomym udziale lub niebiorących udziału w opracowywaniu artykułu;
   - autorstwo grzecznościowe (*gift authorship*) – podawanie jako współautorów osób, których wkład jest oparty jedynie na słabym powiązaniu z badaniem.
3. Podczas zbierania i analizy danych, pisania artykułu i opracowywania elementów graficznych autorzy mogą wspomagać się narzędziami sztucznej inteligencji, ale to oni powinni wnieść główny wkład twórczy w powstanie artykułu i są w pełni odpowiedzialni za treści wygenerowane automatycznie, a tym samym za wszelkie naruszenia etyki publikacyjnej. Są także zobowiązani do poinformowania redakcji o użyciu narzędzi sztucznej inteligencji. Narzędzie takie nie może być wskazane jako współautor. Artykuł całkowicie wygenerowany za pomocą narzędzi sztucznej inteligencji nie może być uznany za oryginalną pracę naukową i zostanie odrzucony.
4. Autorzy są zobowiązani do podania w treści artykułu wszelkich źródeł finansowania badań będących podstawą pracy.
5. Autorzy deklarują w stosownym oświadczeniu, że zgłaszany artykuł nie narusza praw autorskich osób trzecich, nie był dotychczas publikowany i nie został złożony w innym wydawnictwie (także w innej wersji językowej) oraz że jest ich oryginalnym dziełem, i określają swój wkład w opracowanie artykułu. Oświadczają także, że mają zgodę właścicieli materiałów (ikonograficznych itp.) wykorzystanych w pracy na ich opublikowanie (jeśli dotyczy). Podają również informację, czy używali narzędzi sztucznej inteligencji podczas tworzenia

artykułu. W przypadku złożenia przez autorów artykułu w innym wydawnictwie przed ukazaniem się artykułu w „WS" zobowiązują się do niezwłocznego powiadomienia o tym redakcji „WS". Jeżeli doszło do zaprezentowania materiałów, na podstawie których powstał artykuł, np. podczas konferencji, to podczas składania tekstu do publikacji w „WS" autorzy są zobowiązani poinformować o tym redakcję.

6. Autorzy zgłaszający artykuły do publikacji w „WS" biorą udział w procesie recenzji *double-blind peer review*, dokonywanej przez co najmniej dwóch niezależnych ekspertów z danej dziedziny. Po otrzymaniu pozytywnych recenzji autorzy wprowadzają zalecane przez recenzentów poprawki i dostarczają redakcji zaktualizowaną wersję opracowania wraz z pisemnym poświadczeniem uwzględnienia poprawek. Jeśli pojawi się różnica zdań co do zasadności proponowanych zmian, to należy wyjaśnić, które poprawki zostały uwzględnione, a w przypadku ich nieuwzględnienia – uzasadnić swoje stanowisko.
7. Jeżeli autorzy odkryją w swoim maszynopisie lub tekście już opublikowanym błędy, nieścisłości bądź niewłaściwe dane, powinni niezwłocznie poinformować o tym redakcję w celu dokonania korekty, wycofania tekstu lub zamieszczenia sprostowania. W przypadku korekty artykułu już opublikowanego jego nowa wersja jest zamieszczana na stronie internetowej „WS" wraz ze stosownym wyjaśnieniem.

3.2. Odpowiedzialność Rady Naukowej, Kolegium Redakcyjnego i Wydziału Czasopism Naukowych w Departamencie Opracowań Statystycznych GUS

1. Rada Naukowa (RN) kształtuje profil programowy czasopisma, określa kierunki jego rozwoju i konsultuje jego zakres merytoryczny.
2. Kolegium Redakcyjne (KR) podejmuje decyzję o publikacji danego artykułu z uwzględnieniem ocen recenzentów oraz opinii redakcji. W swoich rozstrzygnięciach członkowie KR kierują się kryteriami merytorycznej oceny wartości artykułu, jego oryginalności i jasności przekazu, a także ścisłego związku z celem i zakresem tematycznym „WS". Oceniają artykuły niezależnie od płci, rasy, pochodzenia etnicznego, narodowości, religii, wyznania, światopoglądu, niepełnosprawności, wieku lub orientacji seksualnej ich autorów.
3. Redakcja, wyodrębniona z KR, dokonuje oceny wstępnej nadsyłanych artykułów, która polega na weryfikacji merytorycznej, ocenie ich zgodności z celem i zakresem tematycznym „WS" oraz sprawdzeniu pod względem spełniania wymogów redakcyjnych i przestrzegania zasad rzetelności naukowej. Ponadto redakcja wybiera recenzentów w taki sposób, aby nie wystąpił konflikt interesów, i dba o zapewnienie uczciwego, bezstronnego i terminowego procesu recenzowania.
4. Pracownicy Wydziału Czasopism Naukowych (WCN) w Departamencie Opracowań Statystycznych Głównego Urzędu Statystycznego odpowiadają za sprawny przebieg procesu wydawniczego, politykę informacyjną, zapewnienie anonimowości autorów i recenzentów oraz przygotowanie artykułów do publikacji. Informacje dotyczące artykułu mogą być przekazywane przez WCN wyłącznie autorom, recenzentom, członkom RN i KR oraz wydawcy.
5. W celu uzyskania obiektywnej oceny oryginalności nadsyłanych artykułów przed skierowaniem ich do recenzji pracownicy WCN wykorzystują system antyplagiatowy. W przypadku wykrycia znacznego podobieństwa artykułu do innych prac lub wysokiego prawdopodobieństwa użycia narzędzi sztucznej inteligencji powiadamiają o tym redaktora naczelnego, który razem z KR podejmuje decyzję o przyjęciu lub odrzuceniu artykułu. W przypadku odrzucenia autor otrzymuje decyzję wraz z jej uzasadnieniem.

6. Zmiany dokonane w tekście na etapie przygotowania artykułu do publikacji nie mogą naruszać zasadniczej myśli autorów. Wszelkie modyfikacje o charakterze merytorycznym są z nimi konsultowane.
7. W przypadku podjęcia decyzji o niepublikowaniu artykułu nie może on zostać w żaden sposób wykorzystany przez wydawcę lub uczestników procesu wydawniczego bez pisemnej zgody autorów. Autorzy mogą się odwołać od decyzji o niepublikowaniu artykułu. W tym celu powinni się skontaktować z redaktorem naczelnym lub sekretarzem redakcji „WS" i przedstawić stosowną argumentację. Odwołania autorów są rozpatrywane przez redaktora naczelnego.
8. Członkowie RN i KR ani pracownicy WCN nie mogą pozostawać w jakimkolwiek konflikcie interesów w odniesieniu do artykułów zgłaszanych do publikacji. Przez konflikt interesów należy rozumieć sytuację, w której jakiekolwiek interesy lub zależności (służbowe, finansowe lub inne) mogą mieć wpływ na ocenę artykułu lub decyzję o jego publikacji.
9. Jeżeli na którymkolwiek etapie procesu wydawniczego powstanie uzasadnione podejrzenie, że autorzy dopuścili się nierzetelności naukowej (zob. pkt Odpowiedzialność autorów), redakcja skrupulatnie zbada sprawę ewentualnego nadużycia zgodnie z zasadami COPE określonymi na stronie https://publicationethics.org/guidance/Flowcharts. Jeśli nierzetelność autorów zostanie udowodniona, zgłoszony przez nich artykuł zostanie odrzucony, a autorzy otrzymają informację o podjętej decyzji wraz z uzasadnieniem.
10. W przypadku zgłoszenia przez czytelnika uzasadnionych podejrzeń o nierzetelność naukową autorów opublikowanego artykułu redakcja bada sprawę ewentualnego nadużycia i informuje czytelnika o rezultacie przeprowadzonego postępowania. Jeżeli nadużycie zostanie potwierdzone, to na łamach czasopisma ukaże się stosowna informacja.

3.3. Odpowiedzialność recenzentów

1. Recenzenci przyjmują artykuł do recenzji tylko wtedy, gdy uznają, że:
   - posiadają odpowiednią wiedzę w określonej dziedzinie, aby rzetelnie ocenić pracę;
   - zgodnie z ich stanem wiedzy nie istnieje konflikt interesów w odniesieniu do autorów, przedstawionych w artykule badań i instytucji je finansujących, co potwierdzają w oświadczeniu;
   - mogą wywiązać się z terminu ustalonego przez redakcję, aby nie opóźniać publikacji.
2. Recenzenci wypełniają kartę recenzji, w której oceniają, czy:
   - artykuł może być opublikowany w obecnej formie;
   - artykuł może być opublikowany po uwzględnieniu zalecanych poprawek;
   - artykuł wymaga znacznej modyfikacji i ponownej oceny recenzenta;
   - artykuł nie powinien zostać opublikowany.

   Recenzenci powinni uzasadnić swoją ocenę, przedstawiając stosowną argumentację. Są zobligowani do zachowania obiektywności i poufności oraz powstrzymania się od osobistej krytyki. Niedopuszczalne jest korzystanie z narzędzi sztucznej inteligencji podczas sporządzania recenzji.
3. Recenzenci powinni wskazać ważne dla wyników badań opublikowane prace, które w ich ocenie powinny zostać przywołane w ocenianym artykule.

4. W razie stwierdzenia wysokiego poziomu zbieżności treści recenzowanej pracy z innymi opublikowanymi materiałami lub podejrzenia innych przejawów nierzetelności naukowej recenzenci są zobowiązani poinformować o tym redakcję.
5. Po ukończeniu recenzji przechowywanie przesłanych przez redakcję materiałów (w jakiejkolwiek formie) oraz posługiwanie się nimi przez recenzentów jest niedozwolone.

3.4. Odpowiedzialność wydawcy

1. Materiały opublikowane w „WS" są chronione prawem autorskim. Od 2022 r. autorzy udzielają wydawcy licencji Creative Commons Uznanie autorstwa – Na tych samych warunkach 4.0 (CC BY-SA 4.0). Szczegółowa informacja o prawach autorskich (copyright) jest podawana przy każdym artykule, zarówno w wersji elektronicznej, jak i drukowanej.
2. Wydawca udostępnia pełną treść artykułów w internecie w trybie otwartego dostępu, tj. bezpłatnie i bez technicznych ograniczeń. Użytkownicy mogą czytać, pobierać, kopiować, drukować i wykorzystywać do innych celów artykuły zamieszczone na stronie internetowej czasopisma, zgodnie z zapisami:
   - ustawy o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego w przypadku artykułów zgłoszonych do 31.12.2021 r.;
   - licencji Creative Commons w przypadku artykułów zgłoszonych po 31.12.2021 r.

   Inne sposoby wykorzystania treści artykułów „WS" wymagają zgody wydawcy.
3. Wydawca deklaruje gotowość do opublikowania poprawek, wyjaśnień i przeprosin.

4. Wymogi redakcyjne

Zgodnie z wymogami czasopisma omawiany w artykule problem badawczy powinien być jednoznacznie zdefiniowany oraz istotny dla oceny zjawisk społecznych lub gospodarczych. Artykuł powinien zawierać wyraźnie określony cel badania, precyzyjny opis badanych zjawisk i stosowanych metod, uzyskane wyniki przeprowadzonej analizy oraz autorskie wnioski.

4.1. Struktura i zawartość artykułu

Wymagane elementy artykułu recenzowanego:

1. Tytuł.
2. Dane autora: imię/imiona i nazwisko, afiliacja w języku polskim i angielskim, ORCID, e-mail. W przypadku artykułu wieloautorskiego należy wskazać autora korespondencyjnego.
3. Streszczenie (zalecana objętość – do 1200 znaków ze spacjami, forma bezosobowa). W przypadku artykułu opisującego badanie empiryczne powinno zawierać: cel, przedmiot, okres i metodę badania, źródła danych i najważniejsze wnioski z badania. W przypadku artykułów o innym charakterze należy podać co najmniej cel artykułu, przedmiot i najważniejsze wnioski.

   Streszczenie to podstawowe źródło informacji o artykule, warunkujące też decyzję czytelnika o zapoznaniu się z całą pracą. Dlatego powinno być przygotowane ze szczególną starannością i dbałością o umieszczenie w nim wszystkich wymaganych elementów.

4. Słowa kluczowe – najistotniejsze pojęcia lub wyrażenia użyte w pracy (nie mniej niż trzy). Powinny być zawarte w streszczeniu i/lub tytule.
5. Kod/kody z klasyfikacji Journal of Economic Literature (JEL).
6. Tłumaczenie tytułu, streszczenia i słów kluczowych (na język angielski w przypadku artykułu napisanego w języku polskim, a na język polski w przypadku artykułu napisanego w języku angielskim).
7. W artykule opisującym badanie empiryczne wymagane są następujące części:
   - *Wprowadzenie*, zawierające syntetyczne przedstawienie zagadnień teoretycznych, uzasadnienie podjęcia danego problemu badawczego, cel badania i krytyczne odniesienie do literatury przedmiotu. W wyjątkowych przypadkach, kiedy istotne dla podjętego tematu jest obszerniejsze przedstawienie dyskusji toczącej się w literaturze, przegląd literatury może stanowić odrębną część artykułu;
   - *Metoda badania*, uwzględniająca przedmiot i okres badania, źródła danych i zastosowane metody badawcze, w tym uzasadnienie ich wyboru;
   - *Wyniki badania* – analiza danych oraz interpretacja wyników i odniesienie ich do rezultatów wcześniejszych badań (dyskusja). W uzasadnionych przypadkach dyskusja może stanowić odrębną część artykułu;
   - *Podsumowanie*, które powinno być zwięzłe i odzwierciedlać istotę problemu badawczego przedstawionego w artykule, bez podawania danych liczbowych; końcowe wnioski powinny odnosić się do treści artykułu, a w szczególności do celu badania;
   - *Bibliografia*, zawierająca pełny wykaz prac i materiałów przywołanych w artykule, przygotowana zgodnie z wymogami czasopisma (zob. Przywoływanie źródeł w artykułach napisanych w języku polskim oraz Bibliografia załącznikowa w artykułach napisanych w języku polskim).

   Wszystkie części powinny być opatrzone numerami.
8. Jeżeli podczas gromadzenia i analizy danych, pisania artykułu lub opracowywania elementów graficznych do niego autor korzystał z narzędzi sztucznej inteligencji, to powinien podać w tekście, jakich narzędzi i do czego użył.

   W przypadku artykułu nierecenzowanego nie są wymagane streszczenie, słowa kluczowe ani kody JEL. Bibliografia załącznikowa jest opcjonalna.

4.2. Przygotowanie artykułu

1. Artykuł powinien być utrzymany w formie bezosobowej.
2. Tekst należy zapisać alfabetem łacińskim. Nazwy własne, tytuły itp. oryginalnie zapisane innym alfabetem powinny być poddane transliteracji.
3. Nie należy stosować stylów; formatowanie należy ograniczyć do wymogów redakcyjnych.
4. Objętość artykułu łącznie ze streszczeniem, słowami kluczowymi, bibliografią, tablicami, wykresami i innymi materiałami graficznymi nie powinna być mniejsza niż 10 stron maszynopisu ani przekraczać 20 stron.
5. Edytor tekstu: Microsoft Word, format *.doc lub *.docx.
6. Krój czcionki:
   - Arial – tytuł, autor, streszczenie, słowa kluczowe, kody JEL, śródtytuły, elementy graficzne (tablice, zestawienia, wykresy, schematy), przypisy;
   - Times New Roman – tekst główny, bibliografia.

7. Wielkość czcionki:
   - 14 pkt – tytuł, autor, śródtytuły wyższego rzędu;
   - 12 pkt – tekst główny, śródtytuły niższego rzędu;
   - 10 pkt – pozostałe elementy.
8. Marginesy – 2,5 cm z każdej strony.
9. Interlinia – 1,5 wiersza; tablice i przypisy – 1 wiersz; przed tytułami rozdziałów i podrozdziałów oraz po nich – pusty wiersz.
10. Wcięcie akapitowe – 0,4 cm; bibliografia – bez wcięcia, wysunięcie 0,4 cm.
11. Przy wyliczeniach należy posłużyć się listą punktowaną z punktorami w postaci kropek (wysunięcie 0,4 cm, wcięcie 0 cm); wiersze (oprócz ostatniego) zakończone średnikiem.
12. Strony ponumerowane automatycznie.
13. Tablice i elementy graficzne (wykresy, mapy, schematy) muszą być przywołane w tekście.
14. Wykresy, mapy i schematy należy zamieścić w tekście głównym. Wykresy powinny być edytowalne (optymalnie wykonane w programie Excel; w przypadku wykonania w programie graficznym powinny mieć postać wektorową). Wykresy i inne materiały graficzne należy przekazać osobno, najlepiej w pliku programu Excel lub innym edytowalnym w pakiecie Microsoft Office.
15. Tablice muszą być edytowalne. Nie należy stosować rastrów, cieniowania, pogrubiania czy też podwójnych linii itp.
16. Wskazówki dotyczące opracowywania map znajdują się w publikacji *Mapy statystyczne. Opracowanie i prezentacja danych*, dostępnej na stronie internetowej GUS.
17. Pod tablicami i każdym elementem graficznym należy podać źródło opracowania, a także objaśnić użyte w nich skróty i symbole.
18. Literowe symbole liczb i innych wielkości niezłożonych należy zapisywać małą lub dużą literą i pismem pochyłym (np. a, A, $y(x)$, a_i); wektorów – pismem pochyłym i pogrubionym (np. $\boldsymbol{a}$, $\boldsymbol{A}$, $\boldsymbol{w}$, $\boldsymbol{y}(x)$, $\boldsymbol{w}_i$); macierzy – pismem prostym i pogrubionym (np. $\mathbf{A}$, $\mathbf{a}$, $\mathbf{M}$, $\mathbf{Y}(x)$, $\mathbf{M}_i$).
19. Objaśnienia znaków umownych i zapisów w tablicach: kreska (–) – zjawisko nie wystąpiło; zero (0) – zjawisko istniało w wielkości mniejszej od 0,5; (0,0) – zjawisko istniało w wielkości mniejszej od 0,05; kropka (.) – brak informacji, konieczność zachowania tajemnicy statystycznej, wypełnienie pozycji jest niemożliwe lub niecelowe; „w tym" – oznacza, że nie podaje się wszystkich składników sumy.
20. Stosowane są następujące skróty: tablica – tabl., wykres – wykr.
21. Wszystkie zawarte w artykule informacje, dane i stwierdzenia wykraczające poza wiedzę powszechną – np. wyniki badań innych autorów, zarówno o charakterze empirycznym, jak i koncepcyjnym – muszą być opatrzone przypisem bibliograficznym. Przez wiedzę powszechną należy rozumieć informacje ogólnie znane i niebudzące wątpliwości ani kontrowersji w danej grupie społecznej, np. utworzenie GUS w 1918 r. lub powstanie UE w 1993 r. na podstawie traktatu z Maastricht. Natomiast dane statystyczne udostępniane lub publikowane np. przez GUS lub Eurostat nie należą do takich informacji. Charakteru wiedzy powszechnej nie mają również stwierdzenia odnoszące się do idei, zjawisk i procesów społecznych, politycznych czy gospodarczych. Nawet pozornie zdroworozsądkowe idee zmieniają bowiem swój sens w zależności od kultury, języka lub dyscypliny naukowej, a także bywają w rozmaity sposób konceptualizowane, jak np. pojęcie poznania w naukach społecznych.

    Podanie źródła jest konieczne niezależnie od tego, czy informacje lub stwierdzenia są ujęte w ramy cytatu, czy przedstawione bez dosłownego przytoczenia, np. w formie parafrazy. Jeżeli stwierdzenie może budzić jakiekolwiek wątpliwości odbiorców, autor powinien wskazać stosowne źródło podawanej informacji.

22. Przypisy rzeczowe, słownikowe lub informacyjne należy umieszczać na dole strony. Przypisy bibliograficzne, zgodnie ze standardem APA (American Psychological Association), należy podawać w tekście głównym.
23. Bibliografię należy przygotować zgodnie ze standardem APA.

4.3. Przywoływanie źródeł w artykułach napisanych w języku polskim

4.3.1. Ogólne zasady APA

Wyszczególnienie	Przykład przywołania	
	w odsyłaczu	w treści zdania
Autor indywidualny		
Jeden autor	(Iksiński, 2001)	Iksiński (2001)
Dwóch autorów	(Iksiński i Nowak, 1999)	Iksiński i Nowak (1999)
Trzech autorów lub więcej	(Jankiewicz i in., 2003)	Jankiewicz i in. (2003)
Autor instytucjonalny		
Nazwa funkcjonuje jako powszechnie znany skrótowiec: pierwsze przywołanie w tekście	(International Labour Organization [ILO], 2020)	International Labour Organization (ILO, 2020)
kolejne przywołanie	(ILO, 2020)	ILO (2020)
Pełna nazwa	(Stanford University, 1995)	Stanford University (1995)
Niepełne dane bibliograficzne		
Brak ustalonego autorstwa	(*Skrócony tytuł ...*, 2015)	*Pełny tytuł* (2015)
Brak roku wydania	(Iksiński, b.r.)	Iksiński (b.r.)
Inne przypadki		
Przywoływanie kilku prac (porządek prac – chronologiczny, porządek autorów – alfabetyczny)	(Iksiński, 1997, 1999, 2004a, 2004b; Nowak, 2002)	Iksiński (1997, 1999, 2004a, 2004b) i Nowak (2002)
Przywoływanie publikacji za innym autorem (uwaga: w bibliografii należy wymienić tylko pracę czytaną)	(Nowakowski, 1990, za: Zieniecka, 2007)	Nowakowski (1990, za: Zieniecka, 2007)
Praca tłumaczona, przedruk lub wydanie wznowione	(Adamski, 1857/2020)	Adamski (1857/2020)

Źródło: opracowanie na podstawie: American Psychological Association. (2020). *Publication manual of the American Psychological Association* (wyd. 7). https://doi.org/10.1037/0000165-000.

4.3.2. Szczegółowe wewnętrzne zasady „WS"

4.3.2.1. Adresy portali internetowych, w tym baz danych Głównego Urzędu Statystycznego

Adresy portali internetowych, które są przywoływane w artykule jedynie w celach informacyjnych, należy umieszczać w przypisach dolnych.

W przypadku korzystania z danych pobranych z baz Głównego Urzędu Statystycznego prosimy o podanie w miejscu, w którym baza jest przywoływana po raz pierwszy, pełnej nazwy bazy i jej skrótu (jeśli istnieje), nazwy jej właściciela oraz adresu internetowego w przypisie dolnym. W kolejnych przywołaniach, np. w źródle pod wykresem, należy posługiwać się już tylko pełną lub skróconą nazwą bazy.

Przykłady baz danych GUS	
pierwsze przywołanie	kolejne przywołania
Bank Danych Lokalnych (BDL) Głównego Urzędu Statystycznego + link podany w przypisie dolnym: https://bdl.stat.gov.pl	BDL
Baza Demografia Głównego Urzędu Statystycznego + link podany w przypisie dolnym: https://demografia.stat.gov.pl	Baza Demografia
Dziedzinowe Bazy Wiedzy (DBW) Głównego Urzędu Statystycznego + link podany w przypisie dolnym: https://dbw.stat.gov.pl	DBW

4.3.2.2. Akty prawne

Jeśli autor powołuje się w pracy na akty prawne, powinien za pierwszym razem podać ich pełny oficjalny tytuł; przy kolejnych przywołaniach najczęściej wystarczy nazwa skrócona. W przypadku aktów prawnych zapisanych w innym alfabecie niż łaciński tytuł trzeba poddać transkrypcji. (Informacje dotyczące miejsca publikacji aktu prawnego, takie jak numer dziennika urzędowego, należy podać tylko w opisie zamieszczonym w bibliografii załącznikowej).

Przykłady aktów prawnych	
pierwsze przywołanie	kolejne przywołania
Ustawa z dnia 29 czerwca 1995 r. o statystyce publicznej (dalej: ustawa o statystyce publicznej)	ustawa o statystyce publicznej
Rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 1260/2013 z dnia 20 listopada 2013 r. w sprawie statystyk europejskich w dziedzinie demografii (dalej: rozporządzenie nr 1260/2013)	rozporządzenie nr 1260/2013
Statistics Act	Statistics Act

4.4. Bibliografia załącznikowa w artykułach napisanych w języku polskim

4.4.1. Zasady ogólne

Bibliografia powinna być zamieszczona na końcu opracowania. Opisy bibliograficzne powinny być sporządzone w alfabecie łacińskim.

Źródła należy uszeregować alfabetycznie według nazwiska pierwszego autora, a w przypadku dwóch lub więcej prac tego samego autora – chronologicznie według roku publikacji. Prace bez znanego roku publikacji (oznaczone „b.r.") występują przed pracami ze znanym rokiem publikacji. Jeśli kilka prac tego samego autora zostało opublikowanych w tym samym roku, należy podać je w kolejności alfabetycznej według tytułu i odpowiednio oznaczyć literami a, b, c itd.

Opis bibliograficzny materiałów dostępnych w internecie powinien zawierać link prowadzący do źródłowej strony internetowej lub link DOI. Nie należy podawać linków prowadzących do baz czasopism czy repozytoriów.

4.4.2. Przykłady opisów bibliograficznych

Typ źródła	Przykład opisu bibliograficznego
Artykuł w czasopiśmie	
W wersji: drukowanej	Nazwisko, X. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik* (zeszyt), strona początku–strona końca.
elektronicznej, z DOI	Nazwisko, X., Nazwisko 2, Y. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*(zeszyt), strona początku–strona końca. https://doi.org/xxx.
elektronicznej, bez DOI	Nazwisko, X., Nazwisko 2, Y., Nazwisko 3, Z. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*(zeszyt), strona początku–strona końca. https://xxx.
Opublikowany w trybie online first	Nazwisko, X. (rok). Tytuł artykułu. *Tytuł czasopisma*. Opublikowany w trybie online first. https://xxx.
Artykuł w gazecie codziennej	
W wersji: drukowanej	Nazwisko, X. (rok, dzień i miesiąc). Tytuł artykułu. *Tytuł gazety*, strona lub strona początku–strona końca.
elektronicznej	Nazwisko, X. (rok, dzień i miesiąc). Tytuł artykułu. *Tytuł gazety*. https://xxx. Nazwisko, X. (b.r.). Tytuł artykułu. *Tytuł gazety*. https://xxx. Tytuł artykułu. (rok, miesiąc i dzień). *Tytuł gazety*. https://xxx.
Książka	
W wersji: drukowanej	Nazwisko, X. (rok). *Tytuł książki*. Wydawnictwo.
elektronicznej, z DOI	Nazwisko, X. (rok). *Tytuł książki*. Wydawnictwo. https://doi.org/xxx.
elektronicznej, bez DOI	Nazwisko, X. (rok). *Tytuł książki*. Wydawnictwo. https://xxx.
W przekładzie	Nazwisko, X. (rok). *Tytuł książki* (tłum. Y. Nazwisko). Wydawnictwo.
Wydanie wielotomowe: tom zatytułowany	Nazwisko, X. (rok). *Tytuł książki: numer tomu. Tytuł tomu*. Wydawnictwo.
tom niezatytułowany	Nazwisko, X. (rok). *Tytuł książki* (numer tomu). Wydawnictwo.
Kolejne wydanie	Nazwisko, X. (rok). *Tytuł książki* (numer wydania). Wydawnictwo.
Pod redakcją (niezależnie od języka, w którym książka została wydana)	Nazwisko, X. (red.). (rok). *Tytuł książki*. Wydawnictwo.
Przedruk lub wznowienie	Nazwisko, X. (rok). *Tytuł książki*. Wydawnictwo. (Wydanie pierwotne rok).
W przekładzie	Nazwisko, X. (rok). *Tytuł książki*. (tłum. Y. Nazwisko). Wydawnictwo. (Wydanie pierwotne rok).
Rozdział i hasło słownikowe/encyklopedyczne	
Rozdział w pracy zbiorowej	Nazwisko, X. (rok). Tytuł rozdziału. W: Y. Nazwisko, Z. Nazwisko 2 (red.), *Tytuł książki* (s. strona początku–strona końca). Wydawnictwo. https://doi.org/xxx lub https://xxx.
Hasło ze słownika lub z encyklopedii w wersji: drukowanej	Nazwisko autora hasła, X. (rok). Hasło. W: Y. Nazwisko (red.), *Tytuł*. Wydawnictwo. Hasło. (rok). W: Y. Nazwisko (red.), *Tytuł*. Wydawnictwo.
elektronicznej	Hasło. (rok, dzień i miesiąc lub „b.r."). W: *Tytuł* (np. *Wikipedia* lub *Słownik języka polskiego PWN*). https://xxx.

Typ źródła	Przykład opisu bibliograficznego
Raporty i szara literatura	
Autor: indywidualny	Nazwisko, X. (rok). *Tytuł raportu*. Wydawnictwo. https://doi.org/xxx lub https://xxx.
instytucjonalny	Nazwa instytucji. (rok). *Tytuł raportu.* Wydawnictwo (tylko jeśli wydawcą jest inna instytucja niż instytucja autorska). https://doi.org/xxx lub https://xxx.
Working papers	Nazwisko, X. (rok). *Tytuł pracy* (nazwa serii i numer). https://doi.org/xxx lub https://xxx.
Materiały z konferencji	
Opublikowane jako: druk zwarty	zob. przykład opisu książki lub rozdziału
druk ciągły	zob. przykład opisu artykułu w czasopiśmie
Niepublikowane (jedynie wygłoszone)	Nazwisko, X. (rok, dzień i miesiąc). *Tytuł pracy* [typ wystąpienia, np. referat lub prezentacja]. Nazwa i miejsce (miasto, kraj) konferencji.
Rozprawa doktorska	
Niepublikowana	Nazwisko, X. (rok). *Tytuł pracy* [niepublikowana rozprawa doktorska]. Nazwa instytucji nadającej tytuł doktorski.
Opublikowana, dostępna w internecie	Nazwisko, X. (rok). *Tytuł pracy* [rozprawa doktorska, nazwa instytucji nadającej tytuł doktorski]. https://xxx.
Maszynopis	
Niepublikowany / przygotowywany przez autora / zgłoszony do publikacji, ale jeszcze niezaakceptowany	Nazwisko, X. (rok). *Tytuł* [maszynopis niepublikowany / w przygotowaniu / zgłoszony do publikacji].
Artykuł zaakceptowany do publikacji w czasopiśmie	Nazwisko, X. (w druku). Tytuł artykułu. *Tytuł czasopisma.*
Opublikowany nieformalnie (np. na stronie internetowej autora)	Nazwisko, X., Nazwisko 2, Y. (rok). *Tytuł*. https://xxx.
Akt prawny[a]	
Polski i UE	Pełny tytuł aktu prawnego wraz z numerem/pozycją w dzienniku urzędowym.
Inny	Pełny tytuł aktu prawnego w języku oryginalnym (w przypadku zapisu w innym alfabecie niż łaciński należy podać tylko transkrypcję) wraz z numerem/pozycją w dzienniku urzędowym. https://xxx.
Tekst na stronie internetowej (dostępny tylko online)	
Znana data publikacji, zawartość strony się nie zmienia (jest archiwizowana)	Nazwisko, X. (rok, dzień i miesiąc). *Tytuł*. https://xxx.
Nieznana data publikacji, zawartość strony się zmienia (nie jest archiwizowana)	Nazwa instytucji. (b.r.). *Tytuł*. Pobrane dzień, miesiąc i rok pobrania z https://xxx.

[a] Wewnętrzne zasady „WS".

Typ źródła	Przykład opisu bibliograficznego
Zbiór danych	
Dane opublikowane: znana data publikacji, zawartość zbioru się nie zmienia (jest archiwizowana)	Nazwisko, X. (rok). *Nazwa zbioru danych* [zbiór danych]. Wydawca. https://xxx.
nieznana data publikacji, zawartość zbioru się zmienia (nie jest archiwizowana)	Nazwa instytucji. (b.r.). *Nazwa zbioru danych* [zbiór danych]. Wydawca (tylko jeśli wydawcą jest inna instytucja niż instytucja autorska / właściciel danych). Pobrane dzień, miesiąc i rok pobrania z https://xxx.
Materiały audiowizualne	
Nagranie wideo	Nazwisko, X. (rok, dzień i miesiąc). *Tytuł* [wideo]. Nazwa kanału, na którym nagranie zostało udostępnione (np. YouTube). https://xxx.
Webinar	Nazwisko, X. (rok, dzień i miesiąc). *Tytuł* [webinar]. Nazwa instytucji. https://xxx.
Posty w serwisach społecznościowych	
Post na portalu X lub Instagramie	Nazwisko, X. lub nazwa instytucji [@nazwa użytkownika] (rok, dzień i miesiąc). *Treść – do 20 wyrazów* [post]. Nazwa serwisu społecznościowego (X lub Instagram). https://xxx.
Post na Facebooku	Nazwisko, X. lub nazwa instytucji (rok, dzień i miesiąc). *Treść – do 20 wyrazów* [post]. Facebook. https://xxx. Nazwa instytucji [nazwa użytkownika] (rok, dzień i miesiąc). *Treść – do 20 wyrazów* [post]. Facebook. https://xxx.

Źródło: opracowanie na podstawie: American Psychological Association. (2020). *Publication manual of the American Psychological Association* (wyd. 7). https://doi.org/10.1037/0000165-000.

Praca przygotowana w sposób niezgodny z powyższymi wskazówkami będzie odesłana do autora z prośbą o dostosowanie formy artykułu do wymogów redakcyjnych.

STAŁE DZIAŁY „WS" – ZAKRES TEMATYCZNY
PERMANENT SECTIONS OF *WS* – THEMATIC SCOPE

Tematy artykułów	Topics of the articles
Studia metodologiczne / Methodological studies	
• Oryginalne lub udoskonalone rozwiązania metodologiczne, które mogą znaleźć zastosowanie w analizach statystycznych i służyć podnoszeniu ich jakości • Projektowanie badań statystycznych	• Original or developed methodological solutions which can be applied to statistical analyses and serve to improve their quality • Planning statistical surveys
Statystyka w praktyce / Statistics in practice	
• Nowatorskie zastosowania narzędzi i modeli statystycznych oraz analiza i ocena statystyczna zjawisk społeczno-gospodarczych i innych, prowadzona w szczególności na danych pochodzących z zasobów statystyki publicznej • Wykorzystanie narzędzi informatycznych do uzyskiwania i przetwarzania informacji statystycznych, naliczania i kontroli ujawniania danych oraz prezentacji i rozpowszechniania danych wynikowych	• Innovative applications of statistical tools and models as well as statistical analysis and assessment of social, economic and other phenomena, performed mainly on data produced by official statistics • Application of IT tools to obtain and process statistical information, to calculate data and control the statistical disclosure, and to present and disseminate output data
Studia interdyscyplinarne. Wyzwania badawcze / Interdisciplinary studies. Research challenges	
• Wyzwania badawcze wynikające z rosnących potrzeb użytkowników danych statystycznych i wymagające stosowania rozwiązań z różnych dziedzin nauki • Problematyka wykraczająca poza konwencjonalne tematy związane ze statystyką • Wyniki badań prowadzonych w obrębie różnych dyscyplin z wykorzystaniem metod statystycznych	• Research challenges resulting from growing needs of statistical data users and requiring the application of solutions from various fields of science • Problems beyond the conventional thematic scope related to statistics • Results of research carried out in the framework of several fields of science using statistical methods
Edukacja statystyczna / Statistical education	
• Metody i efekty nauczania statystyki na wszystkich poziomach edukacji • Popularyzacja myślenia statystycznego i rzetelnego posługiwania się informacjami statystycznymi	• Methods and effects of statistical education at all levels of education • Popularisation of statistical thinking and of diligent use of statistical information
Spisy powszechne – problemy i wyzwania / Issues and challenges in census taking	
• Rozwiązania metodologiczne i organizacyjne możliwe do zastosowania podczas przygotowywania i prowadzenia spisów • Praktyczne aspekty związane z gromadzeniem i udostępnianiem danych ze spisów, w tym dotyczące obciążenia odpowiedzi i ochrony tajemnicy statystycznej	• Methodological and organisational solutions which may be implemented in the process of preparing and conducting censuses • Practical aspects of collecting and disseminating census data, including those related to response burden and the protection of statistical confidentiality
Z dziejów statystyki / From the history of statistics	
• Historia prowadzenia obserwacji statystycznych, w tym rozwój metodologii i narzędzi oraz instytucji statystycznych w Polsce i za granicą • Życie i osiągnięcia wybitnych statystyków	• History of statistical observations, including the development of statistical methodologies, tools and institutions in Poland and abroad • Life and achievements of prominent statisticians
In memoriam	
• Nekrologi i artykuły wspomnieniowe o osobach zasłużonych dla statystyki	• Obituaries and articles remembering important people in the world of statistics
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information	
• Dyskusje i polemiki • Sprawozdania z konferencji naukowych • Recenzje książek oraz zestawienia nowości wydawniczych GUS	• Discussions and polemics • Reports from scientific conferences • Book reviews and compilations of Statistics Poland's new publications